Large deviations for bootstrapped empirical measures

José Trashorras* and Olivier Wintenberger†
Université Paris-Dauphine, Ceremade,
Place du Maréchal de Lattre de Tassigny,
75775 Paris Cedex 16 France.

October 21, 2011

Abstract

We investigate the Large Deviations properties of bootstrapped empirical measure with exchangeable weights. Our main result shows in great generality how the resulting rate function combines the LD properties of both the sample weights and the observations. As an application we recover known conditional and unconditional LDPs and obtain some new ones.

Keywords Large deviations, Exchangeable bootstrap

1 Introduction and main results

We say that a sequence of Borel probability measures $(P^n)_{n\geq 1}$ on a topological space $\mathcal{Y}$ obeys a Large Deviation Principle (hereafter abbreviated LDP) with rate function I if I is a non-negative, lower semi-continuous function defined on $\mathcal{Y}$ such that

$$-\inf_{y\in A^o} I(y) \leq \liminf_{n\to\infty} \frac{1}{n}\log P^n(A) \leq \limsup_{n\to\infty} \frac{1}{n}\log P^n(A) \leq -\inf_{y\in \bar{A}} I(y)$$

for any measurable set $A\subset \mathcal{Y}$, whose interior is denoted by A^o and closure by $\bar{A}$. If the level sets $\{y : I(y)\leq \alpha\}$ are compact for every $\alpha<\infty$, I is called a good rate function. With a slight abuse of language we say that a sequence of random variables obeys an LDP when the sequence of measures induced by these random variables obeys an LDP. For a background on the theory of large deviations, see Dembo and Zeitouni [8] and references therein.

*xose@ceremade.dauphine.fr
†wintenberger@ceremade

Our framework is the following: We are given a triangular array $((W_i^n)_{1\le i\le n})_{n\ge1}$ of $\mathbb{R}_+$-valued random variables defined on a probability space $(\Omega, \mathcal{A}, \mathbb{P})$ and such that

(H1) For every $n \ge 1$ we have $\sum_{i=1}^n W_i^n = n$.

(H2) For every $n \ge 1$ the vector $(W_1^n, \dots, W_n^n)$ is n-exchangeable i.e. for every element σ of $\mathfrak{S}_n$ the set of permutations of $\{1, \dots, n\}$ the vectors $(W_1^n, \dots, W_n^n)$ and $(W_{\sigma(1)}^n, \dots, W_{\sigma(n)}^n)$ have the same distribution.

We shall assume that the W_i^n's have some LD properties to be detailed below. We are further given a triangular array $((x_i^n)_{1\le i\le n})_{n\ge1}$ of elements of a Polish space (Σ, d_Σ) such that

$$\mu^n = \frac{1}{n}\sum_{i=1}^n \delta_{x_i^n} \stackrel{w}{\to} \mu \in M_1(\Sigma) \tag{1.1}$$

where $\stackrel{w}{\to}$ stands for weak convergence in the space of Borel probability measures on Σ. Let us recall that $\mu^n \stackrel{w}{\to} \mu$ if and only if for every real-valued, bounded and continuous application f defined on Σ we have $\int_\Sigma f(x)\mu^n(dx) \to \int_\Sigma f(x)\mu(dx)$. Our goal in the present paper is to investigate the LD properties of

$$\mathcal{L}^n = \frac{1}{n}\sum_{i=1}^n W_i^n \delta_{x_i^n}. \tag{1.2}$$

So far LD properties of *families* of randomly weighted empirical measures like (1.2) have been established in the particular case of *independent and identically distributed* $W_1^n, \dots, W_n^n$, see e.g. [15, 7, 23]. Results on the LD properties of some particular cases of $\mathcal{L}^n$ as considered here, i.e. *with exchangeable* $W_1^n, \dots, W_n^n$ are available but their proofs rely on the definition of the chosen sampling weights, see e.g. [2, 7, 6]. Hence, the present paper gives the first derivation of the LD properties of the family of empirical measures $\mathcal{L}^n$ under the natural assumptions **(H1-H2)**. The interest in considering families rather than particular cases lies on the fact that only an upper level of abstraction can reveal the mechanisms that really enter into play.

Here we shall consider $\mathcal{L}^n$ from the bootstrap point of view and adopt its vocabulary [13, 18, 1, 16]. In 1979, in a landmark paper, Efron proposed the following idea: When given a realization $x_1^n, \dots, x_n^n$ of random variables $X_1^n, \dots, X_n^n$ one can easily obtain "additional data" by sampling independent and $\frac{1}{n}\sum_{i=1}^n \delta_{x_i^n}$-distributed random variables $X_1^*, \dots, X_m^*, \dots$. It amounts to sample with replacement from an urn which composition is described by $\frac{1}{n}\sum_{i=1}^n \delta_{x_i^n}$. Often, this is computationally cheap and theoretical studies are available to assess the quality of the distribution of $H(\frac{1}{m}\sum_{i=1}^m \delta_{X_i^*})$ in approximating the distribution of $H(\frac{1}{n}\sum_{i=1}^n \delta_{X_i^n})$ which make it all worthwhile. A rich literature started flourishing on the ground of this idea. It was soon noticed that

the preceding procedure is not the only possible one and that it can be generalized so that $\frac{1}{n}\sum_{i=1}^n W_i^n \delta_{x_i^n}$ under conditions **(H1-H2)** is the right object to be considered. For example, Efron's bootstrap corresponds to $(W_1^n, \dots, W_n^n)$ distributed according to a Multinomial law. The literature on the subject developed into two complementary directions: "conditional" results where $x_1^n, \dots, x_n^n$ are fixed observations filling some conditions and $H(\frac{1}{n}\sum_{i=1}^n W_i^n \delta_{x_i^n})$ is considered and "unconditional" results where the $x_1^n, \dots, x_n^n$ are allowed to fluctuate and $H(\frac{1}{n}\sum_{i=1}^n W_i^n \delta_{X_i^n})$ is considered instead.

Classically the bootstrap scheme is said to be efficient when it mimics the behavior of $H(\frac{1}{n}\sum_{i=1}^n \delta_{X_i^n})$ and one distinguishes between "conditional efficiency" and "unconditional efficiency". For example, Praestgaard and Wellner investigated the Central Limit behavior of $\mathcal{L}^n$ in [24]; they gave a necessary and sufficient condition on the second moments properties of the W_i^n which works for both conditional and unconditional efficiency. Later, Hall and Mammen studied the efficiency of the bootstrap schemes in the Edgeworth expansions at the second order [19]. There, conditions on the fourth order cumulants of the weights are required. In a similar context, Wood already showed in [28] that Efron's bootstrap is efficient to mimic the empirical mean in the moderate deviations regime for observations satisfying the Cramer condition but not for heavier tails.

Following Barbe and Bertail [1], we say that a bootstrap scheme is LD-efficient when the bootstrapped empirical measure has the same LD properties as the original empirical measure. It is a very strong property: thinking of percentile bootstrap's confidence intervals, LD-efficiency says that the relative coverage accuracy tends to 1 exponentially fast. As an application of our general approach, we will be able to discuss both conditional and unconditional LD-efficiency for many classical choices of $((W_i^n)_{1\le i\le n})_{n\ge 1}$ and/or $((X_i^n)_{1\le i\le n})_{n\ge 1}$. Actually we go further in the sense that as an application of our approach we obtain LD results that are new in the literature like e.g. an unconditional LDP for Efron's bootstrap and conditional and unconditional LDP's for iid weighted bootstrap and k-blocks bootstraps.

Now let us describe our results more precisely. To this end we need to introduce some more notations. We consider

$$M_1^1(\mathbb{R}_+) = \left\{ \rho \in M_1(\mathbb{R}_+) \ : \ \int_{\mathbb{R}_+} x \ \rho(dx) = 1 \right\}$$

a subset of

$$\mathcal{W}^1(\mathbb{R}_+) = \left\{ \rho \in M_1(\mathbb{R}_+) \ : \ \int_{\mathbb{R}_+} x \ \rho(dx) < \infty \right\}$$

the so-called Wasserstein space of order 1 on $\mathbb{R}_+$. Both $M_1^1(\mathbb{R}_+)$ and $\mathcal{W}^1(\mathbb{R}_+)$ are endowed with the Wasserstein-1 distance

$$W_{1,|\cdot|}(\rho, \gamma) = \inf_{\pi\in\mathcal{C}(\rho,\gamma)} \left\{ \int_{\mathbb{R}_+\times\mathbb{R}_+} |x-y|\pi(dx, dy) \right\}$$

where $\mathcal{C}(\rho,\gamma)$ is the subset of $M_1(\mathbb{R}_+\times\mathbb{R}_+)$ of couplings of ρ and γ i.e. the set of Borel probability measures π on $\mathbb{R}_+\times\mathbb{R}_+$ such that their first marginal π_1 is ρ and second marginal π_2 is γ. In addition to **(H1-H2)** we shall assume that the $((W_i^n)_{1\le i\le n})_{n\ge 1}$ are such that

(H3) The sequence $(\mathcal{S}^n = \frac{1}{n}\sum_{i=1}^n \delta_{W_i^n})_{n\ge 1}$ satisfies an LDP on $M_1^1(\mathbb{R}_+)$ endowed with the $W_{1,|\cdot|}$ distance with good rate function I^W.

In Section 7 we prove that **(H3)** is valid for a broad collection of sampling weights. Actually we shall see that the LD properties of $(\mathcal{L}^n)_{n\ge 1}$ directly follow from the LD properties of

$$\mathcal{V}^n = \frac{1}{n}\sum_{i=1}^n \delta_{(W_i^n, x_i^n)}$$

on the set

$$\mathcal{M}_1^1(\mathbb{R}_+\times\Sigma) = \left\{\rho\in M_1(\mathbb{R}_+\times\Sigma)\ :\ \int_{\mathbb{R}_+} x\rho_1(dx) = 1\right\}$$

endowed with the distance

$$\Delta(\rho,\gamma) = W_{1,|\cdot|}(\rho_1,\gamma_1) + \beta_{BL,\delta}(\rho,\gamma)$$

where

$$\beta_{BL,\delta}(\rho,\gamma) = \sup_{\substack{f\in C_b(\mathbb{R}_+\times\Sigma)\\ ||f||_\infty + ||f||_{L,\delta}\le 1}} \left\{\left|\int_{\mathbb{R}_+} f(x)\rho(dx) - \int_{\mathbb{R}_+} f(x)\gamma(dx)\right|\right\}$$

is the so-called dual-bounded Lipschitz metric on $M_1(\mathbb{R}_+\times\Sigma)$ and as usual

$$||f||_\infty = \sup_{x\in\mathbb{R}_+\times\Sigma}|f(x)|,\qquad ||f||_{L,\delta} = \sup_{\substack{x,y\in\mathbb{R}_+\times\Sigma\\ x\neq y}}\frac{|f(x)-f(y)|}{\delta(x,y)},$$

δ is a metric on $\mathbb{R}_+\times\Sigma$ compatible with the product topology and $C_b(\mathbb{R}_+\times\Sigma)$ is the set of real-valued, bounded and continuous applications defined on $\mathbb{R}_+\times\Sigma$. For any two probabilities ρ,ν on a measurable space $(E,\mathcal{E})$ we denote by

$$H(\nu|\rho) = \begin{cases} \int_E \mathrm{d}\nu\log\frac{\mathrm{d}\nu}{\mathrm{d}\rho} & \text{if } \nu\ll\rho\\ +\infty & \text{otherwise}\end{cases}$$

the relative entropy of ν with respect to ρ. To any $\rho(dw,dx)\in M_1(\mathbb{R}_+\times\Sigma)$ we associate $\rho_x(dw)\in M_1(\mathbb{R}_+)$ (resp. $\rho_w(dx)\in M_1(\Sigma)$) a stochastic kernel which is the conditional distribution of the first (resp. second) marginal of ρ given the second (resp. first). We summarize this by $\rho(dw,dx) = \rho_x(dw)\otimes\rho_2(dx)$ (resp.

$\rho(dw,dx) = \rho_1(dw) \otimes \rho_w(dx)$). If $\nu, \gamma \in M_1(\mathbb{R}_+ \times \Sigma)$ are such that $\nu_1 = \gamma_1 = \theta$ (resp. $\nu_2 = \gamma_2 = \theta$) then

$$H(\nu|\gamma) = \int_{\mathbb{R}_+} H(\nu_w|\gamma_w)\theta(dw) \tag{1.3}$$

(resp. $H(\nu|\gamma) = \int_\Sigma H(\nu_x|\gamma_x)\theta(dx)$), see Lemma 1.4.3 in [12]. Our first result is the following

Theorem 1.1. *The sequence $(\mathcal{V}^n)_{n\geq 1}$ satisfies an LDP on $\mathcal{M}_1^1(\mathbb{R}_+ \times \Sigma)$ endowed with the distance Δ with good rate function*

$$\mathcal{J}(\rho;\mu) = \begin{cases} H(\rho|\rho_1 \otimes \mu) + I^W(\rho_1) & \text{if } \rho_2 = \mu \\ +\infty & \text{otherwise.} \end{cases}$$

The reason for working on $\mathcal{M}_1^1(\mathbb{R}_+ \times \Sigma)$ endowed with the distance Δ is to make the following map

$$\begin{array}{rccc} F: & \mathcal{M}_1^1(\mathbb{R}_+ \times \Sigma) & \to & M_1(\Sigma) \\ & \rho(dw,dx) & \mapsto & \int_{\mathbb{R}_+} w\rho(dw,dx) \end{array}$$

well-defined and continuous once $M_1(\Sigma)$ is endowed with the weak convergence topology. By contraction (see Theorem 4.2.1 in [8]) an LDP for $(\mathcal{L}^n)_{n\geq 1}$ easily follows from Theorem 1.1

Corollary 1.1. *The sequence $(\mathcal{L}^n)_{n\geq 1}$ satisfies an LDP on $M_1(\Sigma)$ endowed with the weak convergence topology with good rate function*

$$\begin{aligned} \mathcal{K}(\nu;\mu) &= \inf_{\rho: F(\rho)=\nu} \mathcal{J}(\rho;\mu) \\ &= \inf_{\rho_x: F(\rho_x \otimes \mu)=\nu} \left\{ \int_\Sigma H(\rho_x|\rho_1)\mu(dx) + I^W(\rho_1) \right\}. \end{aligned}$$

We shall often see in applications that $I^W(\nu) = H(\nu|\xi)$ for some $\xi \in M_1(\mathbb{R}_+)$ such that

$$\Lambda_\xi(\alpha) = \log \int_{\mathbb{R}_+} e^{\alpha x}\xi(dx) < \infty \tag{1.4}$$

for every $\alpha \in \mathbb{R}$. We denote by

$$\Lambda_\xi^*(x) = \sup_{\alpha\in\mathbb{R}} \{\alpha x - \Lambda_\xi(\alpha)\}. \tag{1.5}$$

the Fenchel-Legendre transform of Λ_ξ. In this particular case we obtain a generalization of Kullback inequality

Lemma 1.1. *If $I^W(\nu) = H(\nu|\xi)$ for some $\xi \in M_1(\mathbb{R}_+)$ then for every $\nu, \mu \in M_1(\Sigma)$ we have*

$$\mathcal{K}(\nu;\mu) \geq \begin{cases} \int_\Sigma \Lambda_\xi^*(\frac{d\nu}{d\mu}(x))\mu(dx) & \text{if } \nu \ll \mu \\ +\infty & \text{otherwise.} \end{cases} \tag{1.6}$$

Moreover, assuming that (1.4) holds for every $\alpha \in \mathbb{R}$, the preceding inequality turns out to be an equality for every $\nu, \mu \in M_1(\Sigma)$ if and only if

$$\lim_{\alpha \to -\infty} \Lambda_\xi^{'}(\alpha) = 0 \text{ and } \lim_{\alpha \to +\infty} \Lambda_\xi^{'}(\alpha) = +\infty. \tag{1.7}$$

Remark that the right hand side of the inequality (1.6) is the rate function obtained in case of iid weights in [7].

Next we allow the x_i^n's to fluctuate and consider a triangular array $((X_i^n)_{1\leq i \leq n})_{n\geq 1}$ of Σ-valued random variables defined on $(\Omega, \mathcal{A}, \mathbb{P})$ such that

(H4) The sequence $(\mathcal{O}^n = \frac{1}{n}\sum_{i=1}^n \delta_{X_i^n})_{n\geq 1}$ satisfies an LDP on $M_1(\Sigma)$ endowed with the weak convergence topology with good rate function I^X.

(H5) For every $n \geq 1$ the vectors $(X_1^n, \dots, X_n^n)$ and $(W_1^n, \dots, W_n^n)$ are independent.

An LDP for

$$V^n = \frac{1}{n}\sum_{i=1}^n \delta_{(W_i^n, X_i^n)}$$

holds as a consequence of Theorem 1.1 and Theorem 2.3 in [17]

Theorem 1.2. *The sequence $(V^n)_{n\geq 1}$ satisfies an LDP on $\mathcal{M}_1^1(\mathbb{R}_+ \times \Sigma)$ endowed with the distance Δ with good rate function*

$$J(\rho) = H(\rho|\rho_1 \otimes \rho_2) + I^W(\rho_1) + I^X(\rho_2). \tag{1.8}$$

Again, by contraction, an LDP for

$$L^n = \frac{1}{n}\sum_{i=1}^n W_i^n \delta_{X_i^n}$$

easily follows from Theorem 1.2

Corollary 1.2. *The sequence $(L^n)_{n\geq 1}$ satisfies an LDP on $M_1(\Sigma)$ endowed with the weak convergence topology with good rate function*

$$\begin{aligned}
K(\nu) &= \inf_{\rho : F(\rho) = \nu} J(\rho) \\
&= \inf_{\rho_2 \in M_1(\Sigma)} \left\{ \inf_{\rho_x : F(\rho_x \otimes \rho_2) = \nu} \left\{ \int_\Sigma H(\rho_x|\rho_1)\rho_2(dx) + I^W(\rho_1) \right\} + I^X(\rho_2) \right\} \\
&= \inf_{\rho_2 \in M_1(\Sigma)} \left\{ \mathcal{K}(\nu; \rho_2) + I^X(\rho_2) \right\}. \qquad (1.9)
\end{aligned}$$

It follows that for every $\nu \in M_1(\Sigma)$ we have $K(\nu) \leq I^X(\nu)$.

The latter inequality illustrates the *smoothing effect* of the random weights $W_1^n, \dots, W_n^n$ on the distribution of L^n. We shall see in most examples that for classical choices of $((W_i^n)_{1 \le i \le n})_{n \ge 1}$ and/or $((X_i^n)_{1 \le i \le n})_{n \ge 1}$ there exists at least one $\nu \in M_1(\Sigma)$ such that $K(\nu) < I^X(\nu)$. Nevertheless, we have the following

Corollary 1.3. *A necessary and sufficient condition on $((W_i^n)_{1 \le i \le n})_{n \ge 1}$ to ensure that for every $((X_i^n)_{1 \le i \le n})_{n \ge 1}$ and every $\nu \in M_1(\Sigma)$ we have $K(\nu) = I^X(\nu)$ is that for every $\nu, \zeta \in M_1(\Sigma)$*

$$\mathcal{K}(\nu; \zeta) = \begin{cases} 0 & \text{if } \nu = \zeta \\ +\infty & \text{otherwise.} \end{cases}$$

It follows from Lemma 1.1 that is condition is satisfied e.g. when the sampling weights are such that $I^W(\nu) = H(\nu|\delta_1)$ as for the delete-$h(n)$ jacknife with $h(n) = o(n)$, see Corollary 2.5 below.

The paper is structured as follows. Section 2 presents several applications of our main results. There we discuss efficiency issues for both conditional and unconditional LDPs. The rest of the paper is devoted to proofs. In particular Section 7 contains the proofs of several sampling weights LDPs. Some of these are new to the literature.

2 Examples of applications

In this section we investigate the LD properties of $(\mathcal{L}^n)_{n \ge 1}$ and $(L^n)_{n \ge 1}$ for several particular choices of $((W_i^n)_{1 \le i \le n})_{n \ge 1}$ and/or $((X_i^n)_{1 \le i \le n})_{n \ge 1}$. To specify our results we need some more notations. For every $\lambda > 0$ we shall denote by $\mathcal{P}(\lambda)$ the Poisson distribution with parameter λ and by $\mathcal{Q}(\lambda)$ the distribution of a random variable Y such that λY is $\mathcal{P}(\lambda)$-distributed. More generally, for every $\lambda, \gamma > 0$ we shall denote by $\mathcal{F}(\lambda, \gamma)$ the distribution of a random variable Y such that λY is $\mathcal{P}(\gamma)$-distributed. For every positive integers m and n and every n-tuple of non-negative numbers $(p_1^n, \dots, p_n^n)$ such that $\sum_{i=1}^n p_i^n = 1$ we shall denote by $\mathrm{Mult}_n(m, (p_1^n, \dots, p_n^n))$ the distribution of $(Y_1, \dots, Y_n)$ the numbers of balls found in n urns labeled $1, \dots, n$ when m balls are thrown in these urns independently, each having probability p_1^n to fall in the urn labeled 1, probability p_2^n to fall in the urn labeled 2, etc...

2.1 Efron's bootstrap and "m out of n" bootstrap

For every $m, n \ge 1$ the weights $(W_1^n, \dots, W_n^n)$ for the "m out of n" bootstrap are defined such that

$$\frac{m}{n}(W_1^n, \dots, W_n^n) \sim \mathrm{Mult}_n(m, (1/n, \dots, 1/n)).$$

Classical Efron's bootstrap corresponds to $m = n$. It emerges from sampling with replacement from the urn containing the observed data. We shall assume

that $m = m(n)$ and that the sequence $(\lambda_n = m(n)/n)_{n\geq 1}$ satisfies $\lim_{n\to\infty} \lambda_n = \lambda > 0$. Quite surprisingly we could not find in the literature a reference for the following, even in the simple $m(n) = n$ case

Theorem 2.1. *The sequence* $(\mathcal{S}^n = \frac{1}{n}\sum_{i=1}^n \delta_{W_i^n})_{n\geq 1}$ *obeys an LDP on* $M_1^1(\mathbb{R}_+)$ *endowed with the* $W_{1,|\cdot|}$ *distance with good rate function* $H(\cdot|\mathcal{Q}(\lambda))$.

First we consider fixed observations $((x_i^n)_{1\leq i\leq n})_{n\geq 1}$ such that (1.1) holds.

Corollary 2.1. *The sequence* $(\mathcal{L}^n)_{n\geq 1}$ *satisfies an LDP on* $M_1(\Sigma)$ *endowed with the weak convergence topology with good rate function*

$$\mathcal{K}(\nu;\mu) = \lambda H(\nu|\mu).$$

By properly rescaling we immediately obtain for every $\lambda > 0$ and every measurable $A \subset M_1(\Sigma)$ that

$$-\inf_{\nu\in A^o} H(\nu|\mu) \leq \liminf_{n\to\infty} \frac{1}{m(n)} \log \mathbb{P}(\mathcal{L}^n \in A) \leq$$
$$\leq \limsup_{n\to\infty} \frac{1}{m(n)} \log \mathbb{P}(\mathcal{L}^n \in A) \leq -\inf_{\nu\in\bar{A}} H(\nu|\mu). \quad (2.1)$$

Next we investigate the LD properties of $(L^n)_{n\geq 1}$ without any other assumption that **(H4-H5)**. To this end we introduce

$$\mathcal{Z} = \left\{\eta \in M_1(\Sigma) \;:\; I^X(\eta) = 0\right\}.$$

It follows from Corollary 1.2 and 2.1 that

Corollary 2.2. *The sequence* $(L^n)_{n\geq 1}$ *satisfies an LDP on* $M_1(\Sigma)$ *endowed with the weak convergence topology with good rate function* K *such that*

$$K(\nu) = \inf_{\zeta\in M_1(\Sigma)} \{\lambda H(\nu|\zeta) + I^X(\zeta)\} \leq \lambda \inf_{\eta\in\mathcal{Z}} H(\nu|\eta).$$

Now we consider some particular cases for $((X_i^n)_{1\leq i\leq n})_{n\geq 1}$. First, we assume that for every $n \geq 1$ the random variables $X_1^n, \ldots, X_n^n$ are independent and identically μ-distributed. Then $\frac{1}{n}\sum_{i=1}^n \delta_{X_i^n} \overset{w}{\to} \mu$ *a.s.* (see Theorem 11.4.1 in [11]) and Corollary 2.1 can be interpreted as a conditional LDP. Hence, any "m out of n" bootstrap such that $\lim_{n\to\infty} m(n)/n = 1$ (in particular Efron's bootstrap) leads to a conditional LDP that coincides with the original LDP in this case. This was first established in [2] for Efron's Bootstrap and in [6] in the general case. Actually the $X_1^n, \ldots, X_n^n$ need not be iid but only that the associated empirical measures satisfy an LDP with rate function $H(\cdot|\mu)$ for Efron's bootstrap to be conditionally LD-efficient. Corollary 2.2 completes the previous result with an unconditional LDP. In this particular case $I^X(\zeta) = H(\zeta|\mu)$ and by taking e.g. $\mu = \frac{9}{10}\delta_0 + \frac{1}{10}\delta_1$, $\nu = \frac{1}{10}\delta_0 + \frac{9}{10}\delta_1$, $\zeta = \frac{1}{2}\delta_0 + \frac{1}{2}\delta_1$ and $\lambda = 1$ we observe that $K(\nu) \leq H(\nu|\zeta) + H(\zeta|\mu) < H(\nu|\mu)$ hence Efron's

bootstrap is not unconditionally LD-efficient. Straightforward use of the same kind of arguments shows that this remark also holds true when $X_1^n, \ldots, X_n^n$ is the result of sampling without replacement from an urn with suitable properties (see Theorem 7.2 in [8] for the reference LDP) or when $X_1^n, \ldots, X_n^n$ are the n first components of an infinitely exchangeable sequence of random variables (see [9] for the reference LDP).

2.2 Iid weighted bootstrap

The weights $(W_1^n, \ldots, W_n^n)$ for an iid-weighted bootstrap are defined on the ground of a sequence $Y_1, \ldots, Y_n, \ldots$ of $\mathbb{R}_+$-valued independent random variables with common distribution ξ. We shall assume that for every $\alpha > 0$ we have $\Lambda_\xi(\alpha) < \infty$ and that $\Lambda_\xi^*(0) = \infty$ (or equivalently $\mathbb{P}(Y_1 = 0) = 0$). The weights $(W_1^n, \ldots, W_n^n)$ are defined by

$$W_1^n = \frac{Y_1}{\frac{1}{n}\sum_{i=1}^n Y_i}, \ldots, W_i^n = \frac{Y_i}{\frac{1}{n}\sum_{i=1}^n Y_i}, \ldots, W_n^n = \frac{Y_n}{\frac{1}{n}\sum_{i=1}^n Y_i}.$$

In order to describe the LD behavior of $(\mathcal{S}^n = \frac{1}{n}\sum_{i=1}^n \delta_{W_i^n})_{n\geq 1}$ we introduce the map

$$\begin{array}{rccc} \mathcal{G}: & \mathcal{W}^1(\mathbb{R}_+) \times \mathbb{R}_+^* & \to & \mathcal{W}^1(\mathbb{R}_+) \\ & (\rho, m) & \mapsto & \mathcal{G}(\rho, m) : A \in \mathcal{B}_{\mathbb{R}_+} \mapsto \rho(mA) \end{array}$$

where for every Borel set $A \in \mathcal{B}_{\mathbb{R}_+}$ and every $m > 0$ we write

$$mA = \{x \in \mathbb{R}_+, \exists y \in A : x = my\}.$$

The continuity of $\mathcal{G}$ is the main argument in the proof of the following

Theorem 2.2. *The sequence $(\mathcal{S}^n = \frac{1}{n}\sum_{i=1}^n \delta_{W_i^n})_{n\geq 1}$ satisfies an LDP on $M_1^1(\mathbb{R}_+)$ endowed with the $W_{1,|\cdot|}$ distance with good rate function*

$$I^W(\rho) = \inf_{m>0} \left\{ H(\mathcal{G}(\frac{1}{m}, \rho)|\xi) \right\}.$$

The previous LDP leads to

Corollary 2.3. *For every $\nu, \mu \in M_1(\Sigma)$ we have*

$$\mathcal{K}(\nu;\mu) \geq \left\{ \begin{array}{ll} \inf_{m>0} \int_\Sigma \Lambda_\xi^*(m\frac{d\nu}{d\mu}(x))\mu(dx) & \textit{if } \nu \ll \mu \\ +\infty & \textit{otherwise.} \end{array} \right.$$

Moreover, the preceding turns out to be an equality for every $\nu, \mu \in M_1(\Sigma)$ if and only if

$$\lim_{\alpha\to-\infty} \Lambda_\xi^{'}(\alpha) = 0 \textit{ and } \lim_{\alpha\to+\infty} \Lambda_\xi^{'}(\alpha) = +\infty.$$

It follows from the previous corollary that there is no distribution ξ such that for every $\nu, \mu \in M_1(\Sigma)$ the identity $\mathcal{K}(\nu;\mu) = H(\nu|\mu)$ holds. Indeed, as soon as there exists $\nu, \mu \in M_1(\Sigma)$ such that $\frac{d\nu}{d\mu}(x) = 0$ on a set A such that $\mu(A) > 0$ one has $\mathcal{K}(\nu;\mu) = \infty$ while it could be possible that $H(\nu|\mu) < \infty$. In words there is no choice of ξ for which one gets a conditional LDP that coincides with the original one for $X_1^n, \dots, X_n^n$ independent and μ-distributed. It is clearly due to the fact that $\Lambda_\xi^*(0) = \infty$ forces all the weights $W_1^n, \dots, W_n^n$ to be positive which is to be compared to e.g. Efron's bootstrap. Finally, as for Efron's bootstrap, one can construct examples to show that in most classical cases the iid-bootstrap is not unconditionally LD-efficient.

2.3 The multivariate hypergeometric bootstrap

Let K be a fixed integer number such that $K \geq 2$. The multivariate hypergeometric bootstrap emerges from the following urn scheme: Put K copies of each observed data in an urn so that the urn contains Kn elements then draw from this urn a sample of size n without replacement. The sampling weights $(W_1^n, \dots, W_n^n)$ take their values in $\{0, 1, \dots, K\}$ under the constraint $\sum_{i=1}^n W_i^n = n$ and are distributed according to

$$\mathbb{P}(W_1^n = w_1^n, \dots, W_n^n = w_n^n) = \frac{C_K^{w_1^n} \cdots C_K^{w_n^n}}{C_{nK}^n}.$$

Let us denote by $\mathfrak{B}(K, \frac{1}{K})$ the Binomial distribution with parameters K and $\frac{1}{K}$.

Theorem 2.3. *The sequence $(\mathcal{S}^n = \frac{1}{n}\sum_{i=1}^n \delta_{W_i^n})_{n\geq 1}$ satisfies an LDP on $M_1^1(\mathbb{R}_+)$ endowed with the $W_{1,|\cdot|}$ distance with good rate function*

$$I^W(\rho) = H(\rho|\mathfrak{B}(K, \frac{1}{K}))$$

Consider fixed observations $((x_i^n)_{1\leq i\leq n})_{n\geq 1}$ such that (1.1) holds, we obtain

Corollary 2.4. *The sequence $(\mathcal{L}^n)_{n\geq 1}$ satisfies an LDP on $M_1(\Sigma)$ endowed with the weak convergence topology with good rate function*

$$\mathcal{K}(\nu;\mu) \geq \begin{cases} \int_\Sigma \Lambda^*_{\mathfrak{B}(K,\frac{1}{K})}(\frac{d\nu}{d\mu}(x))\mu(dx) & \text{if } \nu \ll \mu \\ +\infty & \text{otherwise.} \end{cases}$$

We only obtain an inequality since $\mathfrak{B}(K, \frac{1}{K})$ does not satisfy condition (1.7). Again, there is no integer K such that for every $\nu, \mu \in M_1(\Sigma)$ the identity $\mathcal{K}(\nu;\mu) = H(\nu|\mu)$ holds. Indeed, as soon as there exists $\nu, \mu \in M_1(\Sigma)$ such that $\frac{d\nu}{d\mu}(x) > K$ on a set A such that $\mu(A) > 0$ one has $\mathcal{K}(\nu;\mu) = \infty$ while it could be possible that $H(\nu|\mu) < \infty$. Thus all multivariate hypergeometric bootstraps fail to be conditionally LD-efficients for iid observations. One can construct examples to show that in most classical cases the multivariate hypergeometric bootstrap fails to be unconditionally LD-efficient.

2.4 A bootstrap generated from deterministic weights

The weights for bootstrap schemes defined from deterministic weights are given by

$$(W_1^n,\dots,W_n^n) = (w_{\sigma_n(1)}^n,\dots,w_{\sigma_n(n)}^n)$$

where for every $n \geq 1$ the $w_1^n,\dots,w_n^n$ are fixed non-negative real numbers such that $\sum_{i=1}^n w_i^n = n$ and

$$\frac{1}{n}\sum_{i=1}^n \delta_{w_i^n} \overset{W_1}{\to} \gamma \in M_1^1(\mathbb{R}_+)$$

and σ_n is an uniformly over $\mathfrak{S}_n$ distributed random variable. We clearly have

Theorem 2.4. *The sequence $(\mathcal{S}^n = \frac{1}{n}\sum_{i=1}^n \delta_{W_i^n})_{n\geq 1}$ satisfies an LDP on $M_1^1(\mathbb{R}_+)$ endowed with the $W_{1,|\cdot|}$ distance with good rate function*

$$I^W(\rho) = \begin{cases} 0 & \textit{if } \rho = \gamma \\ +\infty & \textit{otherwise.} \end{cases}$$

An important special case is the grouped, or delete-h jacknife. The grouped jacknife with group block size h may be viewed as a bootstrap generated by permuting the deterministic weights

$$(w_1^n,\dots,w_n^n) \quad = \quad \Bigg(\underbrace{\frac{n}{n-h},\dots,\frac{n}{n-h}}_{n-h}, \quad \underbrace{0,\dots,0}_{h}\Bigg)$$

We shall take $h = h(n)$ such that $\lim_{n\to\infty} h(n)/n = \alpha \in [0,1)$ so

$$\gamma = (1-\alpha)\delta_{\frac{1}{1-\alpha}} + \alpha\delta_0.$$

Corollary 2.5. *If $\alpha > 0$ the sequence $(\mathcal{L}^n)_{n\geq 1}$ satisfies an LDP on $M_1(\Sigma)$ endowed with the weak convergence topology with good rate function*

$$\mathcal{K}(\nu;\mu) = \begin{cases} (1-\alpha)H(\nu|\mu) + \alpha H\left(\frac{\mu-(1-\alpha)\nu}{\alpha}|\mu\right) & \textit{if } \frac{\mu-(1-\alpha)\nu}{\alpha} \in M_1(\Sigma) \\ +\infty & \textit{otherwise.} \end{cases}$$

If $\alpha = 0$ the sequence $(\mathcal{L}^n)_{n\geq 1}$ satisfies an LDP on $M_1(\Sigma)$ endowed with the weak convergence topology with good rate function

$$\mathcal{K}(\nu;\mu) = \begin{cases} 0 & \textit{if } \nu = \mu \\ +\infty & \textit{otherwise.} \end{cases}$$

Naturally this result coincides with Theorem 7.2.1 in [8]. Combining Corollary 1.2 and Corollary 2.5 we obtain an unconditional version of the latter result. To every $\nu \in M_1(\Sigma)$ we associate

$$\mathcal{E}_\nu = \{\zeta \in M_1(\Sigma) \ : \ \frac{\zeta-(1-\alpha)\nu}{\alpha} \in M_1(\Sigma)\}$$

Corollary 2.6. *If $\alpha > 0$ the sequence $(L^n)_{n\geq 1}$ satisfies an LDP on $M_1(\Sigma)$ endowed with the weak convergence topology with good rate function $K(\nu) = \inf_{\zeta\in\mathcal{E}_\nu}\mathcal{U}(\nu,\zeta)$ where*

$$\mathcal{U}(\nu,\zeta) = (1-\alpha)H(\nu|\zeta) + \alpha H\left(\frac{\zeta-(1-\alpha)\nu}{\alpha}|\zeta\right) + I^X(\zeta).$$

If $\alpha = 0$ the sequence $(L^n)_{n\geq 1}$ satisfies an LDP on $M_1(\Sigma)$ endowed with the weak convergence topology with good rate function $K(\nu) = I^X(\nu)$.

2.5 The k-blocks bootstraps

Let us consider the (moving or circular) k-block bootstrap. Consider weights from the "$m = n/k$ out of n" bootstrap:

$$\frac{1}{k}(W_1^n,\dots,W_n^n) \sim \mathrm{Mult}_n(m,(1/n,\dots,1/n)).$$

The k-blocks bootstrapped empirical measure can be written as in [20] with the formula

$$\widetilde{\mathcal{L}}_n = \frac{1}{n}\sum_{i=1}^n W_i^n \frac{1}{k}\sum_{j\sim i}\delta_{x_j^n}$$

where $j\sim i$ means that the j belong to block i. For the moving k-blocks bootstrap, $j\sim i$ if $j\in\{i-k/2,\cdots,i+k/2\}$ miodulo n. With slight modifications we could also consider the circular k-blocks bootstrap where $j\sim i$ if $j\in\{i,\dots,i+k-1\}$ modulo n. Both schemes are asymptotically equivalent as soon as k is fixed as it is the case here. Notice that

$$\widetilde{\mathcal{L}}_n = \frac{1}{n}\sum_{i=1}^n \widetilde{W_i^n}\delta_{x_i^n} \quad \text{with} \quad \widetilde{W_i^n} = \frac{1}{k}\sum_{j\sim i} W_j^n$$

where $(\widetilde{W_1^n},\dots,\widetilde{W_n^n})$ fails to be exchangeable. However, our approach relies on preliminary results like Theorem 1.1 that are general enough to allow us to handle this situation under some mild additional hypothesis. Indeed, assume that the observations $((x_i^n)_{1\leq i\leq n})_{n\geq 1}$ satisfy

$$\frac{1}{n}\sum_{i=1}^n \delta_{(x_i^n,\dots,x_{i+k-1}^n)} \xrightarrow{w} \mu^{(k)} \in M_1(\Sigma^k). \tag{2.2}$$

the i indices being taken modulo n. This situation arises e.g. when we are given the realization $x_1^n,\dots,x_n^n$ of $X_1^n,\dots,X_n^n$ the first n components of a stationary Markov chains $(Y_i)_{i\geq 1}$ with transition probability P and stationary measure μ as in [10] (see also [5]). In this case we get

$$\frac{1}{n}\sum_{i=1}^n \delta_{(X_i^n,\dots,X_{i+k-1}^n)} \xrightarrow{w} \mu\otimes\underbrace{P\otimes\cdots\otimes P}_{k-1} \quad a.s.$$

so (2.2) is satisfied with $\mu^{(k)} = \mu \otimes \underbrace{P \otimes \cdots \otimes P}_{k-1}$. which leads to the following conditional LDP

Theorem 2.5. *The sequence $(\widetilde{\mathcal{L}}_n)$ satisfies an LDP on $M_1(\Sigma)$ endowed the weak convergence topology with good rate function*

$$\widetilde{\mathcal{K}}(\nu; \mu^{(k)}) = \inf\left\{\frac{1}{k}H(\nu^{(k)}|\mu^{(k)}) : \nu^{(k)} \in M_1(\Sigma^k), \frac{1}{k}\sum_{i=1}^{k}\nu_i^{(k)} = \nu\right\}$$

In the particular case of iid observations we further obtain

Corollary 2.7. *The sequence $(\widetilde{\mathcal{L}}_n)$ satisfies an LDP on $M_1(\Sigma)$ endowed the weak convergence topology with good rate function*

$$\widetilde{\mathcal{K}}(\nu; \mu^{(k)}) = H(\nu|\mu).$$

hence the k-blocks bootstrap is conditionnally efficient in this case but fails to be unconditionally efficient for the same reason as Efron's bootstrap, see Section 2.1.

3 Proof of Theorem 1.1

Most of the proof of Theorem 1.1 relies on the proof of a particular case that we describe now: We are given two triangular arrays $((w_i^n)_{1\leq i\leq n})_{n\geq 1}$ and $((x_i^n)_{1\leq i\leq n})_{n\geq 1}$ of elements of $\mathbb{R}_+$ and Σ respectively, possibly with repetition, such that

$$\nu^{1,n} = \frac{1}{n}\sum_{i=1}^{n}\delta_{w_i^n} \overset{W_1}{\to} \nu^1 \in M_1^1(\mathbb{R}_+) \text{ and } \nu^{2,n} = \frac{1}{n}\sum_{i=1}^{n}\delta_{x_i^n} \overset{w}{\to} \nu^2 \in M_1(\Sigma),$$

and such that for every $n \geq 1$ we have $\sum_{i=1}^n w_i^n = n$. Let $(\sigma_n)_{n\geq 1}$ be a sequence of random variables defined on $(\Omega, \mathcal{A}, \mathbb{P})$ such that for every $n \geq 1$ the distribution of σ_n is uniform over $\mathfrak{S}_n$. Let $(T^n)_{n\geq 1}$ be the sequence of random measures defined on $(\Omega, \mathcal{A}, \mathbb{P})$ by

$$T^n = \frac{1}{n}\sum_{i=1}^{n}\delta_{(w^n_{\sigma_n(i)}, x_i^n)} \in \mathcal{M}_1^1(\mathbb{R}_+ \times \Sigma).$$

This is a particular case of $\mathcal{V}^n$. Following the proof of Theorem 1 in [25] we shall prove

Lemma 3.1. *The sequence $(T^n)_{n\geq 1}$ satisfies an LDP on $M_1(\mathbb{R}_+ \times \Sigma)$ endowed with the weak convergence topology with good rate function*

$$I(\rho) = \begin{cases} H(\rho|\nu^1 \otimes \nu^2) & \text{if } \rho_1 = \nu^1 \text{and } \rho_2 = \nu^2 \\ +\infty & \text{otherwise.} \end{cases}$$

A stronger version of Lemma 3.1 i.e. an LDP for $(T^n)_{n\geq 1}$ on $\mathcal{M}_1^1(\mathbb{R}_+\times\Sigma)$ endowed with the distance Δ is proved in Section 3.2. The final proof of Theorem 1 relies on the latter result and Theorem 2.3 in [17] and is given in Section 3.3.

3.1 Proof of Lemma 3.1

Let us introduce some more notations before we carry on with the proof of Lemma 3.1. For every $n\geq 1$ we write

$$\mathcal{P}_n=\left\{\nu\in\mathcal{M}_1^1(\mathbb{R}_+\times\Sigma):\exists\,\sigma\in\mathfrak{S}_n,\ \nu=\frac{1}{n}\sum_{i=1}^n\delta_{(w^n_{\sigma(i)},x^n_i)}\right\}.$$

We introduce two triangular arrays $((L_i^n)_{1\leq i\leq n})_{n\geq 1}$ and $((R_i^n)_{1\leq i\leq n})_{n\geq 1}$ of elements of $\mathbb{R}_+$ and Σ respectively, defined on $(\Omega,\mathcal{A},\mathbb{P})$ and such that for every $n\geq 1$ the $2n$ random variables $L_1^n,\ldots,L_n^n,R_1^n,\ldots,R_n^n$ are mutually independent. We further assume that every L_i^n (resp. R_i^n) is distributed according to $\nu^{1,n}$ (resp. $\nu^{2,n}$). The sequence

$$\mathcal{T}^n=\frac{1}{n}\sum_{i=1}^n\delta_{(L_i^n,R_i^n)}\in M_1(\mathbb{R}_+\times\Sigma)$$

has the following LD behavior

Lemma 3.2. *The sequence $(\mathcal{T}^n)_{n\geq 1}$ satisfies an LDP on $M_1(\mathbb{R}_+\times\Sigma)$ endowed with the weak convergence topology with good rate function $H(\rho|\nu^1\otimes\nu^2)$.*

Proof Since $\nu^{1,n}\overset{W_1}{\to}\nu^1$ we have $\nu^{1,n}\overset{w}{\to}\nu^1$ according to e.g. Theorem 7.11 in [26]. Moreover, since $\nu^{2,n}\overset{w}{\to}\nu^2$ we have $\nu^{1,n}\otimes\nu^{2,n}\overset{w}{\to}\nu^1\otimes\nu^2$ (see [3], Chapter 1, Theorem 3.2). The announced result then follows from Theorem 3 in [2]. $\Box$

Our strategy in proving Lemma 3.1 consists in comparing T^n to random measures coupled to $\mathcal{T}^n$. Comparison is possible because the $\rho\in M_1(\mathbb{R}_+\times\Sigma)$ such that $I(\rho)<+\infty$ can be approached in the weak convergence topology by elements of $\mathcal{P}_n$. Our proof of this property requires to use several metrics on $M_1(\mathbb{R}_+\times\Sigma)$ compatible with the weak convergence topology. This is the reason why in Section 3.1.1 we give a short account on the weak convergence topology prior to the proof of our approximation result. In Section 3.1.2 we construct our coupling. Finally, in Section 3.1.3 we prove the LD bounds of Lemma 3.1.

3.1.1 An approximation result

We are given the Polish space $(\mathbb{R}_+,d_e)$ with $d_e(x,y)=|x-y|$[1]. The distance d_e is not a bounded metric but it is topologically equivalent to the bounded metric

[1] We use this notation to underscore the fact that for Lemma 3.1 to hold it is not necessary for the w_i^n's to be real numbers. It is sufficient to have Polish space valued w_i^n's.

$$\tilde{d}_e(x,y) = \frac{d_e(x,y)}{1+d_e(x,y)}$$

Analogously we define a bounded metric $\tilde{d_\Sigma}$ on Σ on the ground of d_Σ. The product topology on $\mathbb{R}_+ \times \Sigma$ is metrizable by e.g.

$$d_{2,M}((x_1,x_2),(y_1,y_2)) = \max(d_e(x_1,y_1), d_\Sigma(x_2,y_2)) \tag{3.1}$$

or

$$d_{2,+}((x_1,x_2),(y_1,y_2)) = d_e(x_1,y_1) + d_\Sigma(x_2,y_2). \tag{3.2}$$

They both make $\mathbb{R}_+ \times \Sigma$ a Polish space. We can also metrize the product topology on $\mathbb{R}_+ \times \Sigma$ with the analogues $\tilde{d}_{2,M}$ and $\tilde{d}_{2,+}$ of (3.1) and (3.2) built on the ground of $\tilde{d_e}$ and $\tilde{d_\Sigma}$. With a slight abuse of notation we shall denote by

$$\beta_{BL,\delta}(\rho,\nu) = \sup_{\substack{f\in C_b(\mathbb{R}_+) \\ \|f\|_\infty + \|f\|_{L,\delta} \le 1}} \left\{ \left| \int_{\mathbb{R}_+} f d\rho - \int_{\mathbb{R}_+} f d\nu \right| \right\} \tag{3.3}$$

the so-called dual-bounded Lipschitz metric on $M_1(\mathbb{R}_+)$ where δ is either d_e or $\tilde{d_e}$. It is compatible with the weak convergence topology (see [11], Chapter 11, Theorem 11.3.3). According to Kantorovitch-Rubinstein Theorem (see [11], Chapter 11, Theorem 11.8.2) the following metric on $M_1(\mathbb{R}_+)$

$$W_{1,\tilde{d}_e}(\rho,\nu) = \inf_{Q\in\mathcal{C}(\rho,\nu)} \left\{ \int_{\mathbb{R}_+\times\mathbb{R}_+} \tilde{d}_e(x,y) Q(dx,dy) \right\},$$

the so-called Wasserstein-1 metric associated to $\tilde{d_e}$ is compatible with the weak convergence topology as well. However, note that the "analogue" of $W_{1,\tilde{d}_e}$ built on the ground of d_e is not a metric for the weak convergence topology (for an illustration of this fact see [11] p.420-421). Finally we shall denote by

$$\beta_{BL,\tilde{d}_{2,M}}(\rho,\nu) = \sup_{\substack{f\in C_b(\mathbb{R}_+\times\Sigma) \\ \|f\|_\infty + \|f\|_{L,\tilde{d}_{2,M}} \le 1}} \left\{ \left| \int_{\mathbb{R}_+\times\Sigma} f d\rho - \int_{\mathbb{R}_+\times\Sigma} f d\nu \right| \right\} \tag{3.4}$$

and

$$W_{1,\tilde{d}_{2,+}}(\rho,\nu) = \inf_{Q\in\mathcal{C}(\rho,\nu)} \left\{ \int_{(\mathbb{R}_+\times\Sigma)\times(\mathbb{R}_+\times\Sigma)} \tilde{d}_{2,+}(x,y) Q(dx,dy) \right\},$$

two metrics on $M_1(\mathbb{R}_+ \times \Sigma)$ compatible with the weak convergence topology on this set. The following is a key result in the proof of Lemma 3.1.

Lemma 3.3. *Let $\rho \in M_1(\mathbb{R}_+ \times \Sigma)$ be such that $\rho_1 = \nu^1$ and $\rho_2 = \nu^2$. For every $n \geq 1$ there exists a $\rho_n \in \mathcal{P}_n$ such that $\rho_n \xrightarrow{w} \rho$.*

Proof Let $\rho \in M_1(\mathbb{R}_+ \times \Sigma)$ be such that $\rho_1 = \nu^1$ and $\rho_2 = \nu^2$. According to Varadarajan's Lemma (see [11], Chapter 11, Theorem 11.4.1) there exists a family $((u_i^n, v_i^n)_{1\le i\le n})_{n\ge 1}$ of elements of $\mathbb{R}_+ \times \Sigma$ such that

$$\gamma^n = \frac{1}{n}\sum_{i=1}^{n} \delta_{(u_i^n, v_i^n)} \xrightarrow{w} \rho.$$

For every $n \ge 1$ we take $\varphi_n, \tau_n \in \mathfrak{S}_n$ such that

$$\sum_{i=1}^{n} \tilde{d}_e(u_i^n, w_{\varphi_n(i)}^n) = \min_{\varphi\in\mathfrak{S}_n} \left\{ \sum_{i=1}^{n} \tilde{d}_e(u_i^n, w_{\sigma(i)}^n) \right\}$$

and

$$\sum_{i=1}^{n} \tilde{d}_\Sigma(v_i^n, x_{\tau_n(i)}^n) = \min_{\tau\in\mathfrak{S}_n} \left\{ \sum_{i=1}^{n} \tilde{d}_\Sigma(v_i^n, x_{\tau(i)}^n) \right\}.$$

We shall prove that

$$\rho^n = \frac{1}{n}\sum_{i=1}^{n} \delta_{(w_{\varphi_n(i)}^n, x_{\tau_n(i)}^n)} \xrightarrow{w} \rho.$$

To this end it is sufficient to prove that $\beta_{BL,\tilde{d}_{2,M}}(\rho^n, \rho) \to 0$. Let $\varepsilon > 0$ be fixed. There exists an N_0 such that for every $n \ge N_0$ we have

$$\beta_{BL,\tilde{d}_{2,M}}(\rho, \gamma^n) < \varepsilon/3. \tag{3.5}$$

Since $\gamma_1^n \xrightarrow{w} \rho_1 = \nu^1$ there exists an N_1 such that for every $n \ge N_1$ we have $W_{1,\tilde{d}_e}(\gamma_1^n, \nu^{1,n}) < \varepsilon/3$. We will show that due to this for every $n \ge N_1$ we have

$$\frac{1}{n}\sum_{i=1}^{n} \tilde{d}_e(u_i^n, w_{\varphi_n(i)}^n) < \varepsilon/3. \tag{3.6}$$

We shall prove that it leads to

$$\beta_{BL,\tilde{d}_{2,M}}(\hat{\gamma}^n, \gamma^n) < \varepsilon/3 \tag{3.7}$$

for every $n \ge N_1$ where

$$\hat{\gamma}^n = \frac{1}{n}\sum_{i=1}^{n} \delta_{(w_{\varphi_n(i)}^n, v_i^n)}.$$

Analogously to (3.6 - 3.7) one can prove that there exists an N_2 such that for every $n \ge N_2$

$$\beta_{BL,\tilde{d}_{2,M}}(\hat{\gamma}^n, \rho^n) < \varepsilon/3. \tag{3.8}$$

By combining (3.5, 3.7, 3.8) we obtain the announced result.
Proof of (3.6) Since γ_1^n and $\nu^{1,n}$ have finite support every Borel probability

measure Q on $\mathbb{R}_+ \times \mathbb{R}_+$ such that $Q_1 = \gamma_1^n$ and $Q_2 = \nu^{1,n}$ is of the form

$$Q(\alpha) = \frac{1}{n} \sum_{i,j=1}^{n} \alpha_{i,j} \delta_{(u_i^n, w_j^n)} \tag{3.9}$$

where $\alpha = (\alpha_{i,j})_{1 \leq i,j \leq n}$ is an $n \times n$ bi-stochastic matrix. Conversely every $n \times n$ bi-stochastic matrix α defines through (3.9) a Borel probability measure $Q(\alpha)$ on $\mathbb{R}_+ \times \mathbb{R}_+$ such that $Q(\alpha)_1 = \gamma_1^n$ and $Q(\alpha)_2 = \nu^{1,n}$. From the Birkhoff-Von Neumann Theorem we know that every bi-stochastic matrix can be written as a convex combination of permutation matrices. These are $n \times n$ matrices with a single 1 in every line and every column, all other entries being 0 (for a proof of this fact see e.g. [22], Chapter 11, Example 11.2). There is an obvious one-to-one correspondence between elements of $\mathfrak{S}_n$ and $n \times n$ permutation matrices. For every $\delta \in \mathfrak{S}_n$ we shall denote by K_δ the permutation matrix naturally associated to δ by this correspondence. Therefore for any Borel probability measure Q on $\mathbb{R}_+ \times \mathbb{R}_+$ such that $Q_1 = \gamma_1^n$ and $Q_2 = \nu^{1,n}$ i.e. any choice of the components $(\lambda_\delta)_{\delta \in \mathfrak{S}_n}$ of the convex combination $\alpha = \sum_{\delta \in \mathfrak{S}_n} \lambda_\delta K_\delta$ such that $Q = Q(\alpha)$ we have

$$\begin{aligned} \int_{\mathbb{R}_+ \times \mathbb{R}_+} \tilde{d}_e(x,y) Q(dx,dy) &= \sum_{\delta \in \mathfrak{S}_n} \lambda_\delta \Big(\sum_{i=1}^{n} \tilde{d}_e(u_i^n, w_{\delta(i)}^n) \Big) & (3.10) \\ &\geq \frac{1}{n} \sum_{i=1}^{n} \tilde{d}_e(u_i^n, w_{\varphi_n(i)}^n). & (3.11) \end{aligned}$$

Hence for every $n \geq N_1$

$$\begin{aligned} \frac{1}{n} \sum_{i=1}^{n} \tilde{d}_e(u_i^n, w_{\varphi_n(i)}^n) &= W_{1,\tilde{d}_e}(\gamma_1^n, \nu^{1,n}) \\ &< \varepsilon/3 \end{aligned}$$

which proves (3.6).

<u>*Proof of (3.7)*</u> For every $f \in C_b(\mathbb{R}_+ \times \Sigma)$ such that $||f||_\infty + ||f||_{L,\tilde{d}_{2,M}} \leq 1$ we have

$$\begin{aligned} \left| \int_{\mathbb{R}_+ \times \Sigma} f d\hat{\gamma}^n - \int_{\mathbb{R}_+ \times \Sigma} f d\gamma^n \right| &\leq \frac{1}{n} \sum_{i=1}^{n} \left| f(w_{\varphi_n(i)}^n, v_i^n) - f(u_i^n, v_i^n) \right| \\ &\leq \frac{1}{n} \sum_{i=1}^{n} \tilde{d}_{2,M}((w_{\varphi_n(i)}^n, v_i^n), (u_i^n, v_i^n)) \\ &\leq \frac{1}{n} \sum_{i=1}^{n} \tilde{d}_e(u_i^n, w_{\varphi_n(i)}^n) \\ &< \varepsilon/3. \end{aligned}$$

Hence for every $n \geq N_1$ we have $\beta_{BL,\tilde{d}_{2,M}}(\hat{\gamma}^n, \gamma^n) < \varepsilon/3$. $\square$

3.1.2 Coupled empirical measures

To every $n \geq 1$ and every realization of $\mathcal{T}^n$ we associate two elements of $M_1(\mathbb{R}_+ \times \Sigma)$ by

$$W_{1,\tilde{d}_{2,+}}(\mathcal{T}^n, \widetilde{W}^n) = \min_{\nu \in \mathcal{P}_n} \left\{ W_{1,\tilde{d}_{2,+}}(\mathcal{T}^n, \nu) \right\} \tag{3.12}$$

and

$$W_{1,\tilde{d}_{2,+}}(\mathcal{T}^n, \widehat{W}^n) = \max_{\nu \in \mathcal{P}_n} \left\{ W_{1,\tilde{d}_{2,+}}(\mathcal{T}^n, \nu) \right\}. \tag{3.13}$$

In case there are several elements of $\mathcal{P}_n$ achieving the min (resp. the max) $\widetilde{W}^n$ (resp. $\widehat{W}^n$) is picked uniformly at random among these measures.

Lemma 3.4. *For every $n \geq 1$ the random measures $\widetilde{W}^n, \widehat{W}^n$ and T^n are identically distributed over $M_1(\mathbb{R}_+ \times \Sigma)$.*

Proof We shall only prove that $\widetilde{W}^n$ and T^n are identically distributed since the proof with $\widehat{W}^n$ and T^n works the same way. Let $n \geq 1$ be fixed. For the sake of clarity let us first assume that there is no repetition among the $w_1^n, \dots, w_n^n$ and the $x_1^n, \dots, x_n^n$. In this case every $\rho \in \mathcal{P}_n$ corresponds to a single $\tau \in \mathfrak{S}_n$ by

$$\rho = \frac{1}{n} \sum_{i=1}^n \delta_{(w^n_{\tau(i)}, x^n_i)}. \tag{3.14}$$

Next let us consider a fixed realization $(l_i^n, r_i^n)_{1 \leq i \leq n}$ of $(L_i^n, R_i^n)_{1 \leq i \leq n}$ and let us denote by w^n the corresponding value for $\mathcal{T}^n$. Due to the property of the minimizer in the Wasserstein distance between two atomic measures we already employed in the proof of Lemma 3.3, for every $\rho \in \mathcal{P}_n$ (i.e. every $\tau \in \mathfrak{S}_n$ according to (3.14)) there exists a $\sigma \in \mathfrak{S}_n$ such that

$$\begin{aligned} W_{1,\tilde{d}_{2,+}}(\mathrm{w}^n, \rho) &= \frac{1}{n} \sum_{i=1}^n \tilde{d}_{2,+}((l_i^n, r_i^n), (w^n_{\sigma(i)}, x^n_{\sigma \circ \tau(i)})) \\ &= \frac{1}{n} \sum_{i=1}^n \tilde{d}_e(l_i^n, w^n_{\sigma(i)}) + \frac{1}{n} \sum_{i=1}^n \tilde{d}_\Sigma(r_i^n, x^n_{\sigma \circ \tau(i)}). \end{aligned}$$

Thus, since we are looking for the minimum over σ and τ, for a fixed realization $(l_i^n, r_i^n)_{1 \leq i \leq n}$ of $(L_i^n, R_i^n)_{1 \leq i \leq n}$ the corresponding value of $\widetilde{\mathrm{w}}^n$ is obtained by finding η_l and η_r such that

$$\sum_{i=1}^n \tilde{d}_e(l_i^n, w^n_{\eta_l(i)}) = \min_{\sigma \in \mathfrak{S}_n} \left\{ \sum_{i=1}^n \tilde{d}_e(l_i^n, w^n_{\sigma(i)}) \right\}$$

and

$$\sum_{i=1}^{n} \tilde{d_\Sigma}(r_i^n, x_{\eta_r(i)}^n) = \min_{\sigma \in \mathfrak{S}_n} \left\{ \sum_{i=1}^{n} \tilde{d_\Sigma}(r_i^n, x_{\sigma(i)}^n) \right\}$$

and taking

$$\widetilde{\mathrm{w}}^n = \frac{1}{n} \sum_{i=1}^{n} \delta_{(w_{\eta_l(i)}^n, x_{\eta_r(i)}^n)}.$$

In case several η_l and/or η_r realize the minima in the displays above, those defining $\widetilde{\mathrm{w}}^n$ are picked among them uniformly at random. Now, remark that for every $\gamma_l, \gamma_r \in \mathfrak{S}_n$, observing $(l_{\gamma_l(i)}^n, r_{\gamma_r(i)}^n)_{1 \leq i \leq n}$ has the same probability as observing $(l_i^n, r_i^n)_{1 \leq i \leq n}$ and results in $\gamma_l \circ \eta_l$ and $\gamma_r \circ \eta_r$ in defining $\widetilde{\mathrm{w}}^n$ instead of η_l and η_r. Thus, if we consider η_l and η_r as random variables defining $\widetilde{W^n}$, we see that their distribution *conditioned on* W^n is uniform over $\mathfrak{S}_n$. Hence $\widetilde{W^n}$ and T^n are both uniformly distributed over $\mathcal{P}^n$, thus identically distributed. This proof extends easily to the case when there are repetitions among the $w_1^n, \ldots, w_n^n$ or $x_1^n, \ldots, x_n^n$. □

3.1.3 Proof of the LD bounds of Lemma 1

We start the proof of the LD bounds by proving the following

Lemma 3.5. *We have:*

1. *I is a good rate function.*
2. *The sequence $(T^n)_{n \geq 1}$ is exponentially tight.*

Proof
(1) Let $\alpha \geq 0$. We have

$$\begin{aligned} N_\alpha &= \{\rho \in M_1(\mathbb{R}_+ \times \Sigma) : I(\rho) \leq \alpha\} \\ &= \{\rho \in M_1(\mathbb{R}_+ \times \Sigma) : H(\rho | \nu^1 \otimes \nu^2) \leq \alpha\} \cap \{\rho \in M_1(\mathbb{R}_+ \times \Sigma) : \rho_1 = \nu^1 \text{ and } \rho_2 = \nu^2\}. \end{aligned}$$

Thus, for every $\alpha \geq 0$, N_α is the intersection of a compact and a closed subset of $M_1(\mathbb{R}_+ \times \Sigma)$, therefore it is compact.

(2) For every measurable $A \subset M_1(\mathbb{R}_+ \times \Sigma)$ we have

$$\begin{aligned}
\limsup_{n\to\infty}\frac{1}{n}\log\mathbb{P}(T^n\in A^c) &= \limsup_{n\to\infty}\frac{1}{n}\log\mathbb{P}(\mathcal{T}^n\in A^c|\frac{1}{n}\sum_{i=1}^n\delta_{L_i^n}=\nu^{1,n},\frac{1}{n}\sum_{i=1}^n\delta_{R_i^n}=\nu^{2,n})\\
&\le \limsup_{n\to\infty}\frac{1}{n}\log\mathbb{P}(\mathcal{T}^n\in A^c)\\
&\qquad -\liminf_{n\to\infty}\frac{1}{n}\log\mathbb{P}(\frac{1}{n}\sum_{i=1}^n\delta_{L_i^n}=\nu^{1,n})\\
&\qquad -\liminf_{n\to\infty}\frac{1}{n}\log\mathbb{P}(\frac{1}{n}\sum_{i=1}^n\delta_{R_i^n}=\nu^{2,n}).
\end{aligned}$$

Since $(\mathcal{T}^n)_{n\ge1}$ satisfies an LDP on $M_1(\mathbb{R}_+\times\Sigma)$ with a good rate function it is exponentially tight (see [8], Remark a) p.8). Thus for every $\alpha\ge0$ we can chose a compact set $A_\alpha\subset M_1(\mathbb{R}_+\times\Sigma)$ that makes the first term in the last display smaller than $-\alpha-2$. On the other hand it is clear that for every $n\ge1$ we have

$$\mathbb{P}(\frac{1}{n}\sum_{i=1}^n\delta_{L_i^n}=\nu^{1,n})\ge n!\frac{1}{n^n}$$

equality corresponding to the case when there are no ties among the $x_1^n,\dots,x_n^n$. Thus

$$-\liminf_{n\to\infty}\frac{1}{n}\log\mathbb{P}(\frac{1}{n}\sum_{i=1}^n\delta_{L_i^n}=\nu^{1,n})\le1$$

which completes the proof. □

Proof of the lower bound

It is sufficient in order to prove the lower bound of the LDP to prove that

$$-I(\rho)\le\liminf_{n\to\infty}\frac{1}{n}\log\mathbb{P}(T^n\in B(\rho,\varepsilon))$$

holds for every $\rho\in M_1(\mathbb{R}_+\times\Sigma)$ and every $\varepsilon>0$, where $B(\rho,\varepsilon)$ stands for the open ball centered at $\rho\in M_1(\mathbb{R}_+\times\Sigma)$ of radius $\varepsilon>0$ for the $W_{1,\tilde{d}_{2,+}}$ metric. So let $\varepsilon>0$ and $\rho\in M_1(\mathbb{R}_+\times\Sigma)$ be such that $I(\rho)<+\infty$. In particular $\rho_1=\nu^1$ and $\rho_2=\nu^2$. According to Lemma 3.3 there exists a sequence $(\rho^n)_{n\ge1}$ of elements of $M_1(\mathbb{R}_+\times\Sigma)$ such that $\rho^n\in\mathcal{P}_n$ and $\rho^n\xrightarrow{w}\rho$. According to Lemma 3.4

$$\begin{aligned}
\mathbb{P}(T^n\in B(\rho,\varepsilon)) &= \mathbb{P}(\widetilde{W}^n\in B(\rho,\varepsilon))\\
&\ge \mathbb{P}(W_{1,\tilde{d}_{2,+}}(\widetilde{W^n},\mathcal{T}^n)<\frac{\varepsilon}{3},W_{1,\tilde{d}_{2,+}}(\mathcal{T}^n,\rho^n)<\frac{\varepsilon}{3},W_{1,\tilde{d}_{2,+}}(\rho^n,\rho)<\frac{\varepsilon}{6})\\
&\ge \mathbb{P}(W_{1,\tilde{d}_{2,+}}(\rho^n,\mathcal{T}^n)<\frac{\varepsilon}{3},W_{1,\tilde{d}_{2,+}}(\rho^n,\rho)<\frac{\varepsilon}{6})
\end{aligned}$$

since it follows from the definition of $\widetilde{W^n}$ that for every $\rho^n \in \mathcal{V}_n$ we have

$$W_{1,\tilde{d}_{2,+}}(\widetilde{W^n}, \mathcal{T}^n) \leq W_{1,\tilde{d}_{2,+}}(\rho^n, \mathcal{T}^n).$$

On the other hand since $\rho^n \xrightarrow{w} \rho$ we get that for n large enough $\left\{W_{1,\tilde{d}_{2,+}}(\rho^n, \rho) < \frac{\varepsilon}{6}\right\} = \Omega$. Thus, for those n's

$$\begin{aligned}\mathbb{P}(W_{1,\tilde{d}_{2,+}}(\rho^n, \mathcal{T}^n) < \frac{\varepsilon}{3}, W_{1,\tilde{d}_{2,+}}(\rho^n, \rho) < \frac{\varepsilon}{6}) &\geq \mathbb{P}(W_{1,\tilde{d}_{2,+}}(\rho, \mathcal{T}^n) < \frac{\varepsilon}{6}, W_{1,\tilde{d}_{2,+}}(\rho^n, \rho) < \frac{\varepsilon}{6}) \\ &\geq \mathbb{P}(W_{1,\tilde{d}_{2,+}}(\mathcal{T}^n, \rho) < \frac{\varepsilon}{6}).\end{aligned}$$

Finally, it follows from Lemma 3.1 that

$$\begin{aligned}\liminf_{n\to\infty} \frac{1}{n} \log \mathbb{P}(T^n \in B(\rho, \varepsilon)) &\geq \liminf_{n\to\infty} \frac{1}{n} \log \mathbb{P}(W_{1,\tilde{d}_{2,+}}(\rho, \mathcal{T}^n) < \frac{\varepsilon}{6}) \\ &\geq -H(\rho|\nu^1 \otimes \nu^2) = -I(\nu).\end{aligned}$$

Proof of the upper bound

In order to prove the upper bound of the LDP, it is sufficient to prove that it holds for compact subsets of $M_1(\mathbb{R}_+ \times \Sigma)$. Indeed, since $(T^n)_{n\geq 1}$ is an exponentially tight sequence (see Lemma 3.5) the full upper bound will follow from Lemma 1.2.18 in [8]. Let A be a compact subset of $M_1(\mathbb{R}_+ \times \Sigma)$ and let us denote by

$$A_{\nu^1,\nu^2} = \left\{\rho \in A : \rho_1 = \nu^1 \text{ and } \rho_2 = \nu^2\right\}$$

which is a compact subset of $M_1(\mathbb{R}_+ \times \Sigma)$ as well. Since the weak convergence topology on $M_1(\mathbb{R}_+ \times \Sigma)$ is compatible with the $W_{1,\tilde{d}_{2,+}}$ metric, it makes $M_1(\mathbb{R}_+ \times \Sigma)$ a regular topological space: For every $\rho \in A$ such that $\rho \in A^c_{\nu^1,\nu^2}$ there exists $\varepsilon_\rho > 0$ such that $B(\rho, 2\varepsilon_\rho) \cap A_{\nu^1,\nu^2} = \emptyset$. In particular $\bar{B}(\rho, \varepsilon_\rho) \cap A_{\nu^1,\nu^2} = \emptyset$ where $\bar{B}(\rho, \varepsilon)$ denotes the closed ball centered on $\rho \in M_1(\mathbb{R}_+ \times \Sigma)$ of radius $\varepsilon > 0$ for the $W_{1,\tilde{d}_{2,+}}$ metric. On the other hand, since $\rho \mapsto H(\rho|\nu^1 \otimes \nu^2)$ is lower semi-continuous, for every $\rho \in A_{\nu^1,\nu^2}$ and every $\delta > 0$ there exists a $\varphi(\rho, \delta) > 0$ such that

$$\inf_{\gamma \in \bar{B}(\rho, \varphi(\rho,\delta))} H(\gamma|\nu^1 \otimes \nu^2) \geq (H(\rho|\nu^1 \otimes \nu^2) - \delta) \wedge \frac{1}{\delta}.$$

For every $\delta > 0$ we consider the coverage

$$A \subset \left(\cup_{\rho \in A \cap A^c_{\nu^1,\nu^2}} B(\rho, \varepsilon_\rho)\right) \cup \left(\cup_{\rho \in A_{\nu^1,\nu^2}} B(\rho, \frac{\varphi(\rho,\delta)}{8})\right)$$

from which we extract a finite coverage

$$A \subset (\cup_{\rho \in I_1} B(\rho, \varepsilon_\rho)) \cup \left(\cup_{\rho \in I_2} B(\rho, \frac{\varphi(\rho,\delta)}{8})\right)$$

where $I_1 \subset A \cap A^c_{\nu^1,\nu^2}$ and $I_2 \subset A_{\nu^1,\nu^2}$ are finite sets. Then, according to Lemma 1.2.15 in [8]

$$\begin{aligned}
\limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}(T^n \in A) \quad &\leq \quad \max\left\{ \limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}(T^n \in \cup_{\rho\in I_1} \bar{B}(\rho, \varepsilon_\rho) \cap A), \right. \\
&\qquad \left. \limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}(T^n \in \cup_{\rho\in I_2} B(\rho, \frac{\varphi(\rho,\delta)}{8})) \right\}.
\end{aligned}$$

For every $\rho \in I_1$ there can not be an infinite number of integers n_k such that

$$\mathbb{P}(T^{n_k} \in \cup_{\rho\in I_1} \bar{B}(\rho, \varepsilon_\rho) \cap A) \neq 0$$

for otherwise we would get $\bar{B}(\rho, \varepsilon_\rho) \cap A_{\nu^1,\nu^2} \neq \emptyset$. The first term in the max is then equal to $-\infty$. We are left with the second term and according to Lemmas 1,2 and 3 we have

$$\begin{aligned}
\limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}(T^n \in A) \quad &\leq \quad \limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}(T^n \in \cup_{\rho\in I_2} B(\nu, \frac{\varphi(\rho,\delta)}{8})) \\
&\leq \quad \max_{\rho\in I_2} \left\{ \limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}(\widehat{W}^n \in B(\rho, \frac{\varphi(\rho,\delta)}{8})) \right\} \\
&\leq \quad \max_{\rho\in I_2} \left\{ \limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}(W_{1,\tilde{d}_{2,+}}(\rho^n, \widehat{W}^n) < \frac{\varphi(\rho,\delta)}{4}) \right\} \\
&\leq \quad \max_{\rho\in I_2} \left\{ \limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}(W_{1,\tilde{d}_{2,+}}(\rho, W^n) < \frac{\varphi(\rho,\delta)}{2}) \right\} \\
&\leq \quad \max_{\rho\in I_2} \left\{ - \inf_{\gamma\in\bar{B}(\rho,\varphi(\rho,\delta))} H(\gamma|\nu^1 \otimes \nu^2) \right\} \\
&\leq \quad \max_{\rho\in I_2} \left\{ -(H(\rho|\nu^1 \otimes \nu^2) - \delta) \wedge \frac{1}{\delta} \right\} \\
&\leq \quad \max_{\rho\in I_2} \left\{ -(I(\rho) - \delta) \wedge \frac{1}{\delta} \right\} \\
&\leq \quad - \inf_{\rho\in A} \left\{ (I(\rho) - \delta) \wedge \frac{1}{\delta} \right\}.
\end{aligned}$$

By letting $\delta \to 0$ we obtain the announced upper bound, see Remark 1.2.10 in [8].

3.2 A stronger version of Lemma 3.1

In the present section we shall prove the following

Lemma 3.6. *The sequence $(T^n)_{n\geq 1}$ satisfies an LDP on $\mathcal{M}_1^1(\mathbb{R}_+ \times \Sigma)$ endowed with the distance Δ with good rate function I.*

Proof First let us notice that

$$\mathcal{N}_1^1(\mathbb{R}_+ \times \Sigma) = \left\{ \rho \in M_1(\mathbb{R}_+ \times \Sigma) \ : \ \int_{\mathbb{R}_+} x \ \rho_1(dx) \leq 1 \right\}$$

is a closed subset of $M_1(\mathbb{R}_+ \times \Sigma)$ when the latter is endowed with the weak convergence topology. Since for every $n \geq 1$ we have $\mathbb{P}(T_n \in \mathcal{N}_1^1(\mathbb{R}_+ \times \Sigma)) = 1$, Lemma 4.1.5 in [8] implies that $(T_n)_{n\geq 1}$ obeys an LDP on $\mathcal{N}_1^1(\mathbb{R}_+ \times \Sigma)$ endowed with the weak convergence topology, with good rate function I.

Next we prove that the same remains true when $\mathcal{N}_1^1(\mathbb{R}_+ \times \Sigma)$ is endowed with the distance Δ. Indeed, since $(T^n)_{n\geq 1}$ satisfies an LDP on the Polish space $M_1(\mathbb{R}_+ \times \Sigma)$ with a good rate function it is exponentially tight: For every $L > 0$ there exists a compact $A \subset M^1(\mathbb{R}_+ \times \Sigma)$ such that

$$\limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}(T^n \in A^c) \leq -L.$$

The set $A \cap K$ with

$$K = \left\{ \rho \in M^1(\mathbb{R}_+ \times \Sigma) \ : \rho_1 \in \cup_{i=1}^n \{\nu^{1,n}\} \right\}$$

is a compact subset of $\mathcal{N}_1^1(\mathbb{R}_+ \times \Sigma)$ endowed with Δ. Moreover

$$\begin{aligned} \limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}(T^n \in (A\cap K)^c) &\leq \limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}(T^n \in A^c) + \\ &\qquad + \limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}(T^n \in K^c) \\ &\leq -L. \end{aligned}$$

Finally, $\mathcal{M}_1^1(\mathbb{R}_+ \times \Sigma)$ is a closed subset of $(\mathcal{N}_1^1(\mathbb{R}_+ \times \Sigma), \Delta)$ and for every $n \geq 1$ we have $\mathbb{P}(T_n \in \mathcal{M}_1^1(\mathbb{R}_+ \times \Sigma)) = 1$ so again, due to Lemma 4.1.5 in [8] the sequence $(T_n)_{n\geq 1}$ obeys an LDP on $\mathcal{M}_1^1(\mathbb{R}_+ \times \Sigma)$ endowed with the distance Δ with good rate function I. $\square$

3.3 Conclusion of the proof of Theorem 1

In order to conclude the proof of Theorem 1 it is sufficient to establish that the distribution of $\mathcal{V}^n$ on $\mathcal{M}_1^1(\mathbb{R}_+ \times \Sigma)$ is a mixture of Large Deviation Systems (LDS) in the sense of [17]. For the sake of clarity we recover the notations of [17] when identifying the components of the LDS

• $\mathcal{Z} = \mathcal{M}_1^1(\mathbb{R}_+ \times \Sigma)$ is a Polish space when endowed with the distance Δ, see the Appendix.

• $\mathcal{X} = M_1^1(\mathbb{R}_+)$ is a Polish space when endowed with the W_1-distance since it

is a closed subset of the Polish space $(\mathcal{W}^1(\mathbb{R}_+), W_{1,|\cdot|})$ (see e.g. [4]).
• For every $n \geq 1$ we note

$$\mathcal{X}_n = \left\{ \nu \in M_1^1(\mathbb{R}_+) \ : \ \exists\ (w_1, \ldots, w_n) \in (\mathbb{R}_+)^n,\ \nu = \frac{1}{n}\sum_{i=1}^n \delta_{w_i} \right\}$$

and for every $\nu \in \mathcal{X}$ and every $n \geq 1$ there exists a $\nu^n \in \mathcal{X}_n$ such that $\nu^n \overset{W_1}{\to} \nu$, see Lemma A.1 in the Appendix below.
• The map $\pi : \mathcal{Z} \to \mathcal{X}$ defined by $\pi(\nu) = \nu_1$ is continuous and surjective.
• For every $n \geq 1$ and every $\nu = \frac{1}{n}\sum_{i=1}^n \delta_{w_i} \in \mathcal{X}_n$ let P_ν^n be the distribution of $T_n = \frac{1}{n}\sum_{i=1}^n \delta_{(w_{\sigma_n(i)}, x_i^n)}$ under $\mathbb{P}$. The family $\Pi = \{P_\nu^n, \nu \in \mathcal{X}_n, n \geq 1\}$ of finite measures on the Borel σ-field on $\mathcal{Z}$ is such that for every $n \geq 1$ and every $\nu \in \mathcal{X}_n$ we have $P_\nu^n(\pi^{-1}(\{\nu\}^c) = 0$.
• Let Q^n be the distribution of $\frac{1}{n}\sum_{i=1}^n \delta_{W_i^n}$. For every $n \geq 1$ and every and every measurable $A \subset \mathcal{M}_1^1(\mathbb{R}_+ \times \Sigma)$

$$\mathbb{P}(\mathcal{V}^n \in A) = \int_{\mathcal{X}_n} P_\nu^n(A) Q^n(d\nu).$$

All the requirements of Definition 2.1 in [17] are satisfied by our model thanks to Lemma 3.6. It follows from Theorem 2.3 in [17] that the sequence $(\mathcal{V}^n)_{n\geq 1}$ obeys an LDP on $\mathcal{M}_1^1(\mathbb{R}_+ \times \Sigma)$ with distance Δ with good rate function

$$\mathcal{J}(\rho) = \begin{cases} H(\rho|\rho_1 \otimes \mu) + I^W(\rho_1) & \text{if } \ \rho_2 = \mu \\ +\infty & \text{otherwise.} \end{cases}$$

4 Proof of Lemma 1.1

In order to prove (1.6) it is sufficient to consider $\nu, \mu \in M_1(\Sigma)$ such that $\mathcal{K}(\nu; \mu) < \infty$ for otherwise the inequality trivially holds. Then there necessary exists (at least) a $\rho \in \mathcal{M}_1^1(\mathbb{R}_+ \times \Sigma)$ such that $\rho_2 = \mu$ and $F(\rho) = \nu$. For every such ρ the latter reads

$$\nu(A) = \int_{\mathbb{R}_+ \times A} w \rho_x(dw) \mu(dx)$$

for every measurable $A \subset \Sigma$, hence $\nu \ll \mu$ and

$$\frac{d\nu}{d\mu}(x) = \int_{\mathbb{R}_+} w \rho_x(dw) \tag{4.1}$$

μ *a.s.* This shows that $\nu \ll \mu$ is a necessary condition for the existence of ρ such that $\rho_2 = \mu$ and $F(\rho) = \nu$ so, as claimed, the announced inequality holds true as an equality if $\nu \ll \mu$ is not satisfied. Moreover, we have

$$\begin{aligned} H(\rho|\rho_1 \otimes \mu) + H(\rho_1|\xi) &= H(\rho|\xi \otimes \mu) \\ &= \int_\Sigma H(\rho_x|\xi)\mu(dx) \end{aligned}$$

since $\rho_2 = \mu$. It follows from Kullback's inequality (see e.g. Theorem 2.1 in [21]) that

$$H(\rho_x|\xi) \geq \Lambda_\xi^*(\int_{\mathbb{R}_+} w\rho_x(dw))$$

which combined with (4.1) brings the announced inequality. Actually, it shows that a necessary and sufficient condition on ξ to get for every $\nu, \mu \in M_1(\Sigma)$ an equality in (1.6) is that for every $\nu, \mu \in M_1(\Sigma)$ there exists a $\rho \in \mathcal{M}_1^1(\mathbb{R}_+ \times \Sigma)$ such that $\rho_2 = \mu$, $F(\rho) = \nu$ and

$$H(\rho_x|\xi) = \Lambda_\xi^*(\frac{d\nu}{d\mu}(x))$$

μ *a.s.*. According to Kullback's inequality this holds true if and only if there exists $\beta_x \in \mathbb{R}$ such that μ *a.s.* ρ_x can take the form

$$\rho_x(dw) = \frac{1}{Z_x} e^{\beta_x w} \xi(dw) \tag{4.2}$$

and still satisfy

$$\int_{\mathbb{R}_+} w\rho_x(dw) = \frac{d\nu}{d\mu}(x).$$

Let us recall that under (1.4) the map Λ_ξ is C^∞ in $\mathbb{R}$, that for every $\alpha \in \mathbb{R}$

$$\begin{aligned} \Lambda_\xi^{'}(\alpha) &= \frac{\int_{\mathbb{R}_+} we^{\alpha w}\xi(dw)}{\int_{\mathbb{R}_+} e^{\alpha w}\xi(dw)} \\ &= \frac{1}{Z}\int_{\mathbb{R}_+} we^{\alpha w}\xi(dw) \end{aligned}$$

and that $\Lambda_\xi^{''}(\alpha) > 0$ for every $\alpha \in \mathbb{R}$, see Section 2.2.1 in [8]. So finally it appears that

$$\lim_{\alpha \to -\infty} \Lambda_\xi^{'}(\alpha) = 0 \text{ and } \lim_{\alpha \to +\infty} \Lambda_\xi^{'}(\alpha) = +\infty.$$

is a necessary and sufficient condition to get for every $\nu, \mu \in M_1(\Sigma)$ and every $x \in \Sigma$ such that $\frac{d\nu}{d\mu}(x) > 0$ a ρ_x of the form (4.2) with $\int_{\mathbb{R}_+} w\rho_x(dw) = \frac{d\nu}{d\mu}(x)$ and $H(\rho_x|\xi) = \Lambda_\xi^*(\frac{d\nu}{d\mu}(x))$. If $\frac{d\nu}{d\mu}(x) = 0$ then one takes $\rho_x = \delta_0$ and still gets $\int_{\mathbb{R}_+} w\rho_x(dw) = \frac{d\nu}{d\mu}(x)$ and $H(\rho_x|\xi) = \Lambda_\xi^*(\frac{d\nu}{d\mu}(x)) = \Lambda_\xi^*(0)$. The announced result follows.

5 Proof of Theorem 1.2

In order to prove Theorem 1.2 it is sufficient to establish that the distribution of V^n on $\mathcal{M}_1^1(\mathbb{R}_+ \times \Sigma)$ is a mixture of LDS. Again we recover the notations of [17] when identifying the components of the LDS

- $\mathcal{Z} = \mathcal{M}_1^1(\mathbb{R}_+ \times \Sigma)$ is a Polish space when endowed with the distance Δ.
- $\mathcal{X} = M_1(\Sigma)$ is a Polish space when endowed with the weak convergence

topology.

• For every $n \geq 1$ we note

$$\mathcal{X}_n = \left\{ \nu \in M_1^1(\Sigma) \ : \ \exists \ (x_1, \ldots, x_n) \in \Sigma^n, \ \nu = \frac{1}{n} \sum_{i=1}^{n} \delta_{x_i} \right\}$$

and according to Varadarajan's Lemma for every $\nu \in \mathcal{X}$ and every $n \geq 1$ there exists a $\nu^n \in \mathcal{X}_n$ such that $\nu^n \stackrel{w}{\to} \nu$.

• The map $\pi : \mathcal{Z} \to \mathcal{X}$ defined by $\pi(\nu) = \nu_2$ is continuous and surjective.

• For every $n \geq 1$ and every $\nu = \frac{1}{n}\sum_{i=1}^n \delta_{x_i} \in \mathcal{X}_n$ let P_ν^n be the distribution of $T_n = \frac{1}{n}\sum_{i=1}^n \delta_{W_i^n, x_i}$ under $\mathbb{P}$. The family $\Pi = \{P_\nu^n, \nu \in \mathcal{X}_n, n \geq 1\}$ of finite measures on the Borel σ-field on $\mathcal{Z}$ is such that for every $n \geq 1$ and every $\nu \in \mathcal{X}_n$ we have $P_\nu^n(\pi^{-1}(\{\nu\}^c) = 0$.

• Let Q^n be the distribution of $\frac{1}{n}\sum_{i=1}^n \delta_{X_i^n}$. For every $n \geq 1$ and every and every measurable $A \subset \mathcal{M}_1^1(\mathbb{R}_+ \times \Sigma)$

$$\mathbb{P}(\mathcal{V}^n \in A) = \int_{\mathcal{X}_n} P_\nu^n(A) Q^n(d\nu).$$

All the requirements of Definition 2.1 in [17] are satisfied by our model thanks to Theorem 1.1. It follows from Theorem 2.3 in [17] that the sequence $(V^n)_{n\geq 1}$ obeys an LDP on $\mathcal{M}_1^1(\mathbb{R}_+ \times \Sigma)$ with distance Δ with good rate function

$$J(\rho) = H(\rho | \rho_1 \otimes \rho_2) + I^W(\rho_1) + I^X(\rho_2).$$

6 Proof of Corollary 1.3

In view of (1.9) a necessary condition on $((W_i^n)_{1\leq i\leq n})_{n\geq 1}$ for $K = I^X$ is that for every $\nu, \zeta \in M_1(\Sigma)$ and every $((X_i^n)_{1\leq i\leq n})_{n\geq 1}$

$$I^X(\nu) - I^X(\zeta) \leq \mathcal{K}(\nu, \zeta).$$

If we consider $X_1^n, \ldots, X_n^n$ resulting from sampling without replacement on an urn which composition $x_1^n, \ldots, x_n^n$ satisfies $\frac{1}{n}\sum_{i=1}^n \delta_{x_i^n} \stackrel{w}{\to} \zeta$ we know that $\frac{1}{n}\sum_{i=1}^n \delta_{X_i^n}$ satisfies an LDP with good rate function

$$I^X(\theta) = \begin{cases} 0 & \text{if } \theta = \zeta \\ \infty & \text{otherwise.} \end{cases}$$

so the announced condition is necessary. It is also clearly sufficient and the announced result follows.

7 Sample weights Large Deviations Principles

In this section we prove the results given in Section 2.

7.1 Efron's bootstrap and "m out of n" bootstrap

7.1.1 Proof of Theorem 2.1

The proof relies on the combination of a coupling construction and a Sanov's result in a strong topology.

Coupling Poisson and Multinomial distributions.

Let $m, n \geq 1$ and $Z_1^n, \dots, Z_n^n$ be independent random variables such that every Z_i^n is $\mathcal{P}(m/n)$-distributed. We shall interpret each Z_i^n as the number of balls randomly thrown in an urn labeled i. To $(Z_1^n, \dots, Z_n^n)$ we couple $(M_1^n, \dots, M_n^n)$ in the following way
• If $\sum_{i=1}^n Z_i^n = m$ we take $M_i^n = Z_i^n$ for every $1 \leq i \leq n$.
• If $\sum_{i=1}^n Z_i^n > m$ we pick uniformly at random $\sum_{i=1}^n Z_i^n - m$ balls in the urns. We define $(M_1^n, \dots, M_n^n)$ as the new occupation numbers of the urns.
• If $\sum_{i=1}^n Z_i^n < m$ we add $m - \sum_{i=1}^n Z_i^n$ balls into the urns. The urns are chosen independently for each added ball with probability $1/n$. Again, we denote by $(M_1^n, \dots, M_n^n)$ the new occupation numbers of the urns.

Lemma 7.1. *The vector $(M_1^n, \dots, M_n^n)$ is $Mult_n(m, (1/n, \dots, 1/n))$-distributed.*

The coupling is optimal in the following sense:

Lemma 7.2. *For every non-negative real numbers $x_1, \dots, x_n$ such that $\sum_{i=1}^n x_i = n$ we have*

$$W_{1,|\cdot|}\Big(\frac{1}{n}\sum_{i=1}^n \delta_{\frac{n}{m}M_i^n}, \frac{1}{n}\sum_{i=1}^n \delta_{\frac{n}{m}Z_i^n}\Big) \leq W_{1,|\cdot|}\Big(\frac{1}{n}\sum_{i=1}^n \delta_{x_i}, \frac{1}{n}\sum_{i=1}^n \delta_{\frac{n}{m}Z_i^n}\Big).$$

Proof of Lemma 7.1 Remark that the conditional law of $(Z_1, \dots, Z_n)$ given $\sum_{i=1}^n Z_i^n = k$ is $\mathrm{Mult}_n(k, (1/n, \dots, 1/n))$. Thus, if $k = m$ the result is obvious. If $k < m$ or $k > m$, the coupling is done such that we move from the law of k balls randomly putted in n urns to m balls into the same urns by picking or putting at random. We finally get the $\mathrm{Mult}_n(m, (1/n, \dots, 1/n))$ law. $\square$

Proof of Lemma 7.2 By definition

$$\begin{aligned} W_{1,|\cdot|}\Big(\frac{1}{n}\sum_{i=1}^n \delta_{\frac{n}{m}M_i^n}, \frac{1}{n}\sum_{i=1}^n \delta_{\frac{n}{m}Z_i^n}\Big) &= \sup_{\substack{f\in C_b(\mathbb{R}_+)\\ ||f||_{L,|\cdot|}\leq 1}} \left\{ \left| \frac{1}{n}\sum_{i=1}^n f\Big(\frac{n}{m}M_i^n\Big) - \frac{1}{n}\sum_{i=1}^n f\Big(\frac{n}{m}Z_i^n\Big) \right| \right\} \\ &\leq \frac{1}{m}\sum_{i=1}^n |M_i^n - Z_i^n| \,. \end{aligned}$$

Due to the construction of $(M_1^n, \dots, M_n^n)$ we have that either all $M_i^n - Z_i^n \leq 0$ or all $M_i^n - Z_i^n \geq 0$, so

$$
\begin{aligned}
\frac{1}{m}\sum_{i=1}^{n}|M_i^n - Z_i^n| &= \left|\frac{1}{m}\sum_{i=1}^{n}(M_i^n - Z_i^n)\right| \\
&= \left|1 - \frac{1}{m}\sum_{i=1}^{n} Z_i^n\right|.
\end{aligned}
$$

Moreover, for every non-negative real numbers $x_1, \ldots, x_n$ such that $\sum_{i=1}^n x_i = n$ we have

$$
\begin{aligned}
W_{1,|\cdot|}(\frac{1}{n}\sum_{i=1}^{n}\delta_{x_i}, \frac{1}{n}\sum_{i=1}^{n}\delta_{\frac{n}{m}Z_i^n}) &= \sup_{\substack{f\in C_b(\mathbb{R}_+)\\ ||f||_{L,|\cdot|}\leq 1}} \left\{\left|\frac{1}{n}\sum_{i=1}^{n} f(x_i) - \frac{1}{n}\sum_{i=1}^{n} f(\frac{n}{m}Z_i^n)\right|\right\} \\
&\geq \left|\frac{1}{n}\sum_{i=1}^{n}(x_i - \frac{n}{m}Z_i^n)\right| \\
&= \left|1 - \frac{1}{m}\sum_{i=1}^{n} Z_i^n\right|,
\end{aligned}
$$

which concludes the proof. □

A Sanov's result in a strong topology.

Let $((Z_1^n, \ldots, Z_n^n)_{1\leq i\leq n})_{n\geq 1}$ be a triangular array of random variables such that for every $n \geq 1$ the $Z_1^n, \ldots, Z_n^n$ are independent and identically $\mathcal{P}(\lambda_n)$-distributed and such that $\lim_{n\to\infty} \lambda_n = \lambda > 0$.

Lemma 7.3. *The sequence $(\mathcal{R}^n = \frac{1}{n}\sum_{i=1}^n \delta_{\frac{1}{\lambda_n}Z_i^n})_{n\geq 1}$ obeys an LDP on $\mathcal{W}^1(\mathbb{R}_+)$ endowed with the $W_{1,|\cdot|}$ distance with good rate function $H(\cdot|\mathcal{Q}(\lambda))$.*

Proof of Lemma 7.3 According to Theorem 3 in [2] the sequence $(\mathcal{R}^n)_{n\geq 1}$ obeys an LDP on $M_1(\mathbb{R}_+)$ when the latter is endowed with the weak convergence topology with good rate function $H(\cdot|\mathcal{Q}(\lambda))$. Moreover, for every $n \geq 1$ we have $\mathcal{R}^n \in \mathcal{W}^1(\mathbb{R}_+)$ and, since all the exponential moments of $\mathcal{Q}(\lambda)$ are finite, for every $\nu \in M_1(\mathbb{R}_+)$ if $H(\nu|\mathcal{Q}(\lambda) < \infty$ then necessarily $\nu \in \mathcal{W}^1(\mathbb{R}_+)$, see e.g. Lemma 2.1 in [14]. Thus $(\mathcal{R}^n)_{n\geq 1}$ obeys an LDP on $\mathcal{W}^1(\mathbb{R}_+)$ endowed with the weak convergence topology with good rate function $H(\cdot|\mathcal{Q}(\lambda))$, see Lemma 4.1.5 in [8]. In particular it is exponentially tight i.e. for every $L > 0$ there exists a compact $K_L \subset \mathcal{W}^1(\mathbb{R}_+)$ such that

$$
\limsup_{n\to\infty} \frac{1}{n}\log \mathbb{P}(\mathcal{R}^n \notin K_L) < -L.
$$

To strengthen the topology for which the LDP holds it is sufficient to prove that the sequence $(\mathcal{R}^n)_{n\geq 1}$ is exponentially tight in the stronger $W_{1,|\cdot|}$ topology, see

Corollary 4.2.6 in [8]. To this end we follow the proof of Theorem 1.1. in [27]. For every $L > 0$ we introduce

$$A_L = \left\{ \nu \in \mathcal{W}^1(\mathbb{R}_+) : \int_{\mathbb{R}_+} \Lambda^*_{\mathcal{Q}(\lambda)}(x)\nu(dx) \leq L \right\}.$$

We prove that for every $L > 0$ the set $C_L = K_L \cap A_L$ is $W_{1,|\cdot|}$-compact. First notice that A_L is closed for the weak convergence topology hence C_L is compact for this topology. Thus, to prove that C_L is $W_{1,|\cdot|}$-compact it remains to prove that C_L has uniformly integrable first moments. Let $S(N) = \Lambda^*_{\mathcal{Q}(\lambda)}(N)/N$. According to Lemma 2.2.20 in [8] we have $\lim_{N\to\infty} S(N) = \infty$. For every $\nu \in C_L$ we have

$$\begin{aligned}
\int_{x\geq N} x\nu(dx) &\leq \frac{1}{S(N)} \int_{x\geq N} \frac{\Lambda^*_{\mathcal{Q}(\lambda)}(x)}{x} x\nu(dx) \\
&\leq \frac{1}{S(N)} \int_{\mathbb{R}_+} \Lambda^*_{\mathcal{Q}(\lambda)}(x)\nu(dx) \\
&\leq \frac{L}{S(N)}
\end{aligned}$$

hence C_L is $W_{1,|\cdot|}$-compact. Since

$$\limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}(\mathcal{R}^n \notin C_L) \leq \max\left(-L, \limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}(\mathcal{R}^n \notin A_L)\right) \tag{7.1}$$

we are led to consider $\mathbb{P}(\mathcal{R}^n \notin A_L)$. According to Chebichev's inequality for every $\delta \in (0, 1)$

$$\begin{aligned}
\mathbb{P}(\mathcal{R}^n \notin A_L) &= \mathbb{P}(\frac{1}{n}\sum_{i=1}^n \Lambda^*_{\mathcal{Q}(\lambda)}(Z_i^n) > L) \\
&\leq \exp(-\delta n L) \left(\mathbb{E} \exp(\delta \Lambda^*_{\mathcal{Q}(\lambda)}(Z_1^n))\right)^n .
\end{aligned} \tag{7.2}$$

But for every $x \in \mathbb{R}$ and every $\lambda > 0$ we have $\Lambda^*_{\mathcal{Q}(\lambda)}(x) = \Lambda^*_{\mathcal{P}(\lambda)}(\lambda x)$ hence

$$\Lambda^*_{\mathcal{Q}(\lambda)}(x) = \begin{cases} \lambda - \lambda x + \lambda x \log x & \text{if } x \geq 0 \\ \infty & \text{otherwise.} \end{cases} \tag{7.3}$$

So we get for every $x \in \mathbb{R}$ the relationship $\Lambda^*_{\mathcal{Q}(\lambda)}(x) = (\lambda/\lambda_n)\Lambda^*_{\mathcal{Q}(\lambda_n)}(x)$ so

$$\mathbb{P}(\mathcal{R}^n \notin A_L) \leq \exp(-\delta n L) \left(\mathbb{E} \exp(\delta(\lambda/\lambda_n)\Lambda^*_{\mathcal{Q}(\lambda)}(Z_1^n))\right)^n .$$

According to Lemma 5.1.14 in [8] for n large enough we have

$$\mathbb{E} \exp(\delta(\lambda/\lambda_n)\Lambda^*_{\mathcal{Q}(\lambda)}(Z_1^n)) \leq \frac{2}{1 - \delta(\lambda/\lambda_n)}$$

which combined with (7.1) and (7.2) shows that $(\mathcal{R}^n)_{n\geq 1}$ is exponentially tight in the $W_{1,|\cdot|}$ topology. The announced result follows. $\square$

Conclusion of the proof of Lemma 2.1

LD upper bound Let F be a $W_{1,|\cdot|}$-closed subset of $M_1^1(\mathbb{R}_+)$. Since $M_1^1(\mathbb{R}_+)$ is a closed subset of $\mathcal{W}^1(\mathbb{R}_+)$, F is also a closed subset of $\mathcal{W}^1(\mathbb{R}_+)$. For every $n \geq 1$ let $Z_1^n, \ldots, Z_n^n$ be independent and identically $\mathcal{P}(\lambda_n)$ distributed random variables and $(M_1^n, \ldots, M_n^n)$ the vector associated to $Z_1^n, \ldots, Z_n^n$ by the coupling construction described in Section 7.1.1. According to Lemma 7.1 $(M_1^n, \ldots, M_n^n)$ is $\mathrm{Mult}_n(m(n), (1/n, \ldots, 1/n))$-distributed, which is also the distribution of $Z_1^n, \ldots, Z_n^n$ conditioned on $\{\sum_{i=1}^n Z_i^n = m(n)\}$. According to Lemma 7.3

$$\begin{aligned}
\limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}\Big(\frac{1}{n}\sum_{i=1}^n \delta_{W_i^n} \in F\Big) &= \limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}\Big(\frac{1}{n}\sum_{i=1}^n \delta_{\frac{n}{m(n)}M_i^n} \in F\Big) \\
&\leq \limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}\Big(\frac{1}{n}\sum_{i=1}^n \delta_{\frac{1}{\lambda_n}Z_i^n} \in F\Big) \\
&\qquad - \liminf_{n\to\infty} \frac{1}{n} \log \mathbb{P}\Big(\sum_{i=1}^n Z_i^n = m(n)\Big) \\
&\leq -\inf_{\nu\in F} H(\nu|\mathcal{Q}(\lambda))
\end{aligned}$$

since $\liminf_{n\to\infty} \frac{1}{n} \log \mathbb{P}(\sum_{i=1}^n Z_i^n = m(n)) = 0$.

LD lower bound Let O be an open subset of $M_1^1(\mathbb{R}_+)$. In order to prove the LD lower bound it is sufficient to prove that for every $\rho \in O$ we have

$$\liminf_{n\to\infty} \frac{1}{n} \log \mathbb{P}\Big(\frac{1}{n}\sum_{i=1}^n \delta_{W_i^n} \in O\Big) \geq -H(\rho|\mathcal{Q}(\lambda)).$$

So let $\rho \in O$ and $\varepsilon > 0$ be small enough to ensure that $B(\rho, \varepsilon) \subset O$. We can assume that $H(\nu|\mathcal{Q}(\lambda)) < \infty$ for otherwise the LD lower bound trivially holds. According to Lemma A.1 proved in the Appendix for every $\rho \in M_1^1(\mathbb{R}_+)$ there exists a triangular array of non-negative real numbers $((x_i^n)_{1\leq i\leq n})_{n\geq 1}$ such that $\rho^n = \frac{1}{n}\sum_{i=1}^n \delta_{x_i^n} \overset{W_1}{\to} \rho$ and such that for every $n \geq 1$ we have $\sum_{i=1}^n x_i^n = n$.

Again, let $Z_1^n, \ldots, Z_n^n$ be independent and identically $\mathcal{P}(\lambda_n)$ distributed random variables and $(M_1^n, \ldots, M_n^n)$ the vector associated to $Z_1^n, \ldots, Z_n^n$ by the coupling construction described above. In particular

$$W_{1,|\cdot|}\Big(\frac{1}{n}\sum_{i=1}^n \delta_{\frac{n}{m(n)}M_i^n}, \frac{1}{n}\sum_{i=1}^n \delta_{\frac{1}{\lambda_n}Z_i^n}\Big) \leq W_{1,|\cdot|}\Big(\frac{1}{n}\sum_{i=1}^n \delta_{\frac{1}{\lambda_n}Z_i^n}, \frac{1}{n}\sum_{i=1}^n \delta_{x_i^n}\Big).$$

So,

$$\begin{aligned}
\mathbb{P}(\frac{1}{n}\sum_{i=1}^{n}\delta_{W_i^n}\in O) &= \mathbb{P}(\frac{1}{n}\sum_{i=1}^{n}\delta_{\frac{n}{m(n)}M_i^n}\in O)\\
&\geq \mathbb{P}(W_{1,|\cdot|}(\frac{1}{n}\sum_{i=1}^{n}\delta_{\frac{n}{m(n)}M_i^n},\frac{1}{n}\sum_{i=1}^{n}\delta_{\frac{1}{\lambda_n}Z_i^n})<\varepsilon/3,\\
&\quad W_{1,|\cdot|}(\rho^n,\frac{1}{n}\sum_{i=1}^{n}\delta_{\frac{1}{\lambda_n}Z_i^n})<\varepsilon/3,\ W_{1,|\cdot|}(\rho^n,\rho)<\varepsilon/3)\\
&\geq \mathbb{P}(W_{1,|\cdot|}(\rho^n,\frac{1}{n}\sum_{i=1}^{n}\delta_{\frac{1}{\lambda_n}Z_i^n})<\varepsilon/3,\ W_{1,|\cdot|}(\rho^n,\rho)<\varepsilon/3)\\
&\geq \mathbb{P}(W_{1,|\cdot|}(\rho,\frac{1}{n}\sum_{i=1}^{n}\delta_{\frac{1}{\lambda_n}Z_i^n})<\varepsilon/6)
\end{aligned}$$

for every n large enough and the LD lower bound follows from Lemma 7.3. □

7.1.2 Proof of Corollary 2.1

First notice that $\Lambda_{\mathcal{Q}(\lambda)}(\alpha)=\lambda(e^{\alpha/\lambda}-1)$ so condition (1.7) is satisfied. Using (7.3) we immediately obtain the announced result. □

□

7.1.3 Proof of Corollary 2.2

Let us prove the inequality in Corollary 2.2. Let $\nu\in M_1(\Sigma)$ be such that $\inf_{\eta\in\mathcal{Z}}H(\nu|\eta)<\infty$. Necessarily for every $\eta\in\mathcal{Z}$ such that $H(\nu|\eta)<\infty$ we have $\nu\ll\eta$. Now for every such η consider $\rho\in\mathcal{M}_1^1(\mathbb{R}_+\times\Sigma)$ defined by $\rho_2=\eta$ while the regular conditional distribution of its first marginal given the second is

$$\rho_x(dw)=\begin{cases}\mathcal{F}(\lambda,\lambda\frac{d\nu}{d\eta}(x))(dw) & \text{if } \frac{d\nu}{d\eta}(x)>0\\ \delta_0(dw) & \text{otherwise.}\end{cases}$$

Let us check that $F(\rho)=\nu$. Indeed, every measurable $A\subset\Sigma$ can be decomposed into $A=A_\nu\cup A_\nu^\perp$ where $A_\nu=A\cap\mathrm{Support}(\nu)$ and

$$\begin{aligned}
F(\rho)(A) &= \int_{\mathbb{R}_+\times A} w\rho_x(dw)\eta(dx)\\
&= \int_{A_\nu}\left(\int_{\mathbb{R}_+} w\mathcal{F}(\lambda,\lambda\frac{d\nu}{d\eta}(x))(dw)\right)\eta(dx)+\int_{A_\nu^\perp}\left(\int_{\mathbb{R}_+} w\delta_0(dw)\right)\eta(dx)\\
&= \int_{A_\nu}\frac{d\nu}{d\eta}(x)\eta(dx)\\
&= \nu(A_\nu)=\nu(A).
\end{aligned}$$

Moreover by taking $\Sigma_1 = \Sigma \cap \{\frac{d\nu}{d\eta}(x) > 0\}$ and $\Sigma_2 = \Sigma \cap \{\frac{d\nu}{d\eta}(x) = 0\}$ we get

$$\begin{aligned}
H(\rho|\rho_1 \otimes \eta) + H(\rho_1|\mathcal{Q}(\lambda)) + I^X(\eta) &= H(\rho|\mathcal{Q}(\lambda) \otimes \eta) \\
&= \int_\Sigma H(\rho_x(\cdot)|\mathcal{Q}(\lambda))\eta(dx) \\
&= \int_{\Sigma_1} H(\mathcal{P}(\lambda \frac{d\nu}{d\eta}(x))|\mathcal{P}(\lambda))\eta(dx) + \\
&\qquad + \int_{\Sigma_2} H(\delta_0|\mathcal{P}(\lambda))\eta(dx) \\
&= \int_{\Sigma_1} \lambda(1 - \frac{d\nu}{d\eta}(x) + \frac{d\nu}{d\eta}(x) \log \frac{d\nu}{d\eta}(x))\eta(dx) + \lambda\eta(\Sigma_2) \\
&= \lambda H(\nu|\eta),
\end{aligned}$$

hence $K(\nu) \leq \lambda \inf_{\eta \in \mathcal{Z}} H(\nu|\eta)$. □

7.2 Iid weighted bootstrap

The following

Lemma 7.4. *$\mathcal{G}$ is continuous.*

is the main argument in the proof of Theorem 2.2

7.2.1 Proof of Lemma 7.4

We consider a sequence of probability measures $(\rho^n)_{n\geq 1}$ such that $\rho^n \overset{W_1}{\to} \rho$ and a sequence of positive numbers $m_n \to m > 0$ and prove that $\mathcal{G}(\rho^n, m_n) \overset{W_1}{\to} \mathcal{G}(\rho, m)$. It is not difficult to show that for every $f \in C_b(\mathbb{R}_+^*)$, every $\rho \in \mathcal{W}^1(\mathbb{R}_+)$ and every $m > 0$ we have

$$\int_{\mathbb{R}_+} f(x)\mathcal{G}(\rho, m)(dx) = \int_{\mathbb{R}_+} f(\frac{x}{m})\rho(dx).$$

So $\int_{\mathbb{R}_+} x\mathcal{G}(\rho^n, m_n)(dx) = \int_{\mathbb{R}_+} \frac{x}{m_n}\rho^n(dx) \to \frac{1}{m}\int_{\mathbb{R}_+} x\rho(dx) = \int_{\mathbb{R}_+} x\mathcal{G}(\rho, m)(dx)$. Hence, according to Theorem 7.11 in [26] we are left to prove that for every $f \in C_b(\mathbb{R}_+^*)$ we have

$$\int_{\mathbb{R}_+} f(x)\mathcal{G}(\rho^n, m_n)(dx) \to \int_{\mathbb{R}_+} f(x)\mathcal{G}(\rho, m)(dx).$$

Since $\rho^n \overset{W_1}{\to} \rho$ we have $\rho^n \overset{w}{\to} \rho$ hence for every $\varepsilon > 0$ there exists a compact

$K_\varepsilon \subset \mathbb{R}_+^*$ such that $\sup_{n\geq 1}\{\rho^n(K_\varepsilon^c), \rho(K_\varepsilon^c)\} < \varepsilon$. Thus

$$\begin{aligned}
&\left|\int_{\mathbb{R}_+} f(x)\mathcal{G}(\rho^n, m_n)(dx) - \int_{\mathbb{R}_+} f(x)\mathcal{G}(\rho, m)(dx)\right| \\
= \;& \left|\int_{\mathbb{R}_+} f(\frac{x}{m_n})\rho^n(dx) - \int_{\mathbb{R}_+} f(\frac{x}{m})\rho(dx)\right| \\
\leq \;& \left|\int_{K_\varepsilon} f(\frac{x}{m_n})\rho^n(dx) - \int_{K_\varepsilon} f(\frac{x}{m})\rho(dx)\right| + 2\varepsilon||f||_\infty \\
\leq \;& \int_{K_\varepsilon} \left|f(\frac{x}{m_n}) - f(\frac{x}{m})\right| \rho^n(dx) + \\
& \qquad + \left|\int_{K_\varepsilon} f(\frac{x}{m})\rho^n(dx) - \int_K f(\frac{x}{m})\rho(dx)\right| + 2\varepsilon||f||_\infty \qquad (7.4)
\end{aligned}$$

Since f is continuous it is uniformly continuous over compact subsets of $\mathbb{R}_+^*$ so there is an N_0 such that for every $n \geq N_0$

$$\sup_{x\in K_\varepsilon} \left|f(\frac{x}{m_n}) - f(\frac{x}{m})\right| < \varepsilon.$$

By taking N_0 large enough we can make the second term in (7.4) as small as desired, which concludes the proof. □

7.2.2 Proof of Theorem 2.2

According to [27], that all the exponential moments of ξ are finite is a necessary and sufficient condition for $\frac{1}{n}\sum_{i=1}^n \delta_{Y_i}$ to satisfy an LDP on $\mathcal{W}^1(\mathbb{R}_+)$ endowed with $W_{1,|\cdot|}$ with good rate function $H(\,\cdot\,|\mu)$. Next, since

$$\begin{array}{cccc} G: & \mathcal{W}^1(\mathbb{R}_+) & \to & \mathbb{R}_+ \\ & \nu & \mapsto & \int_{\mathbb{R}_+} x\nu(dx) \end{array}$$

is continuous when $\mathcal{W}^1(\mathbb{R}_+)$ is furnished with the W_1 distance we obtain that $(\frac{1}{n}\sum_{i=1}^n \delta_{Y_i}, \frac{1}{n}\sum_{i=1}^n Y_i)$ satisfies an LDP on the product space $\mathcal{W}^1(\mathbb{R}_+) \times \mathbb{R}_+$ with good rate function

$$R(\rho, x) = \begin{cases} H(\rho|\mu_1) & \text{if } \int_{\mathbb{R}_+} u\rho(du) = x \\ +\infty & \text{otherwise.} \end{cases}$$

Since $\Lambda^*(0) = \infty$ we have $R(\rho, 0) = \infty$. Hence, according to Lemma 4.1.5 in [8], an LDP for $(\frac{1}{n}\sum_{i=1}^n \delta_{Y_i}, \frac{1}{n}\sum_{i=1}^n Y_i)_{n\geq 1}$ holds on $\mathcal{W}^1(\mathbb{R}_+) \times \mathbb{R}_+^*$ with the same rate function R. Since $\mathcal{G}$ is continuous and

$$\frac{1}{n}\sum_{i=1}^n \delta_{W_i^n} = \mathcal{G}(\frac{1}{n}\sum_{i=1}^n \delta_{Y_i}, \frac{1}{n}\sum_{i=1}^n Y_i)$$

$(\mathcal{S}^n)_{n\geq 1}$ obeys an LDP on $\mathcal{W}^1(\mathbb{R}_+)$ due to the contraction principle, see Theorem 4.2.1 in [8]. Finally, the announced result follows from Lemma 4.1.5 in [8] since for every $n \geq 1$ $\mathcal{S}^n \in M_1^1(\mathbb{R}_+)$ and the latter is a closed subset of $\mathcal{W}^1(\mathbb{R}_+)$. □

7.2.3 Proof of Corollary 2.3

Let $\nu \in M_1(\Sigma)$ be such that $\nu \ll \mu$ for otherwise we already know that $\mathcal{K}(\nu;\mu) = +\infty$. We have

$$\begin{aligned}
\mathcal{K}(\nu;\mu) &= \inf_{\rho_x : F(\rho_x \otimes \mu) = \nu} \left\{ H(\rho|\rho_1 \otimes \mu) + \inf_{m>0} \left\{ H(\mathcal{G}(\frac{1}{m}, \rho_1)|\xi) \right\} \right\} \\
&= \inf_{m>0} \inf_{\rho_x : F(\rho_x \otimes \mu) = \nu} \{ H(\rho|\rho_1 \otimes \mu) + H(\rho_1|\mathcal{G}(m,\xi) \} \\
&= \inf_{m>0} \inf_{\rho_x : F(\rho_x \otimes \mu) = \nu} \{ H(\rho|\mathcal{G}(m,\xi) \otimes \mu) \} \\
&= \inf_{m>0} \inf_{\rho_x : F(\rho_x \otimes \mu) = \nu} \left\{ \int_\Sigma H(\rho_x|\mathcal{G}(m,\xi))\mu(dx) \right\} \\
&\geq \inf_{m>0} \int_\Sigma \Lambda_\xi^*(m \frac{d\nu}{d\mu}(x))\mu(dx)
\end{aligned}$$

where, to establish the last inequality and characterize the equality, we proceed as in the proof of Lemma 1.1. □

7.3 The multivariate hypergeometric bootstrap

Since the W_i^n's take values in the finite set $\{0,\dots,K\}$ the proof of this LDP closely follows the proof of Sanov's Theorem for finite alphabets as exposed in Section 2.1.1 in [8] . This is the reason why we shall not give all the details of the proof. The main step is to notice that for every $(w_1^n,\dots,w_n^n) \in \{0,\dots,K\}^n$ such that $\sum_{i=1}^n w_i^n = n$ if we introduce

$$(\nu^n(1),\dots,\nu^n(K)) = (\frac{1}{n}\sum_{i=1}^n \delta_{w_i^n}(1),\dots,\frac{1}{n}\sum_{i=1}^n \delta_{w_i^n}(K))$$

we get $\sum_{k=1}^K k\nu^n(k) = 1$ and

$$\mathbb{P}(W_1^n = w_1^n,\dots,W_n^n = w_n^n) = R(n)\Pi_{k=1}^K \left(C_K^k (\frac{1}{K})^k (1-\frac{1}{K})^{K-k} \right)^{n\nu^n(k)} \tag{7.5}$$

with $\frac{1}{n}\log R(n) \to 0$. For every integer $n \geq 1$, every $\nu \in M_1^1(\{0,\dots,K\})$ of the form $(\frac{k_1}{n},\dots,\frac{k_K}{n})$ with $k_1,\dots,k_K$ integers there exists $T_n(\nu)$ vectors $(w_1^n,\dots,w_n^n) \in \{0,\dots,K\}^n$ such that $\nu = \frac{1}{n}\sum_{i=1}^n \delta_{w_i^n}$ with

$$(n+1)^{-K} e^{-nH(\nu)} \leq T_n(\nu) \leq e^{-nH(\nu)} \tag{7.6}$$

where $H(\nu) = -\sum_{i=1}^{K} \nu(i) \log \nu(i)$, see Lemma 2.1.8 in [8]. Hence, combining (7.5) and (7.6) for every integer $n \geq 1$, every $\nu \in M_1^1(\{0, \ldots, K\})$ of the form $(\frac{k_1}{n}, \ldots, \frac{k_K}{n})$ with $k_1, \ldots, k_K$ integers we get

$$R_1(n) e^{-nH(\nu|\mathfrak{B}(K,\frac{1}{K}))} \leq \mathbb{P}(\mathcal{S}^n = \nu) \leq R_2(n) e^{-nH(\nu|\mathfrak{B}(K,\frac{1}{K}))}$$

with $\frac{1}{n} \log R_{1,2}(n) \to 0$. The proof then follows as in Theorem 2.1.10 in [8] until its conclusion. $\square$

7.4 A bootstrap generated from deterministic weights

7.4.1 Proof of Corollary 2.5

First we consider $\alpha > 0$. Let $\nu \in M_1(\Sigma)$ be such that $\frac{\mu-(1-\alpha)\nu}{\alpha} \in M_1(\Sigma)$. Then $\rho \in \mathcal{M}_1^1(\mathbb{R}_+ \times \Sigma)$ defined by $\rho_1 = \gamma$ while the regular conditional distribution of its second marginal given the first is

$$\rho_{\frac{1}{1-\alpha}}(dx) = \nu(dx) \text{ and } \rho_0(dx) = \frac{\mu(dx) - (1-\alpha)\nu(dx)}{\alpha}$$

is the only element of $\mathcal{M}_1^1(\mathbb{R}_+ \times \Sigma)$ that satisfies $F(\rho) = \nu, \rho_1 = \gamma$ and $\rho_2 = \mu$. Since for this particular ρ we have

$$H(\rho|\rho_1 \otimes \rho_2) = (1-\alpha) H(\nu|\mu) + \alpha H\left(\frac{\mu - (1-\alpha)\nu}{\alpha} | \mu\right)$$

we obtain an upper-bound on $\mathcal{K}$ as announced. To prove the reverse inequality let $\nu \in M_1(\Sigma)$ be such that $\mathcal{K}(\nu;\mu) < \infty$ for otherwise the announced result trivially holds. Necessarily there exists a $\rho \in \mathcal{M}_1^1(\mathbb{R}_+ \times \Sigma)$ such that $F(\rho) = \nu, \rho_1 = \gamma$ and $\rho_2 = \mu$. These conditions are only met by the probability measure ρ introduced above. In particular $\frac{\mu-(1-\alpha)\nu}{\alpha}$ must be a probability and the reverse inequality holds.

If $\alpha = 0$ then $\gamma = \delta_1$ so the only $\nu \in M_1(\Sigma)$ such that there exists $\rho \in \mathcal{M}_1^1(\mathbb{R}_+ \times \Sigma)$ such that $F(\rho) = \nu$ and $\rho_2 = \mu$ is μ and necessarily $\rho = \delta_1 \otimes \mu$. The announced result follows. $\square$

7.4.2 Proof of Corollary 2.6

First we consider $\alpha > 0$. Let $\nu \in M_1(\Sigma)$ be such that $\inf_{\zeta \in \mathcal{E}_\nu} \mathcal{U}(\nu, \zeta) < \infty$ holds. Then for every $\zeta \in \mathcal{E}_\nu$ we have $\frac{\zeta-(1-\alpha)\nu}{\alpha} \in M_1(\Sigma)$ and $\rho \in \mathcal{M}_1^1(\mathbb{R}_+ \times \Sigma)$ defined by $\rho_1 = \gamma$ and

$$\rho_{\frac{1}{1-\alpha}}(dx) = \nu(dx) \text{ and } \rho_0(dx) = \frac{\zeta(dx) - (1-\alpha)\nu(dx)}{\alpha}$$

satisfies $F(\rho) = \nu, \rho_1 = \gamma$ and $\rho_2 = \zeta$ hence $\mathcal{K}(\nu;\mu) \leq \mathcal{U}(\nu, \zeta)$ and the announced upper bound on $\mathcal{K}$ follows.To prove the reverse inequality let us assume that $\mathcal{K}(\nu;\mu) < \infty$. Then there exists a $\rho \in \mathcal{M}_1^1(\mathbb{R}_+ \times \Sigma)$ such that $F(\rho) = \nu$

and $\rho_1 = \gamma$. Necessarily $\rho_{\frac{1}{1-\alpha}}(dx) = \nu(dx)$ and ρ_2 is such that $\rho_0(dx) = \frac{\zeta(dx)-(1-\alpha)\nu(dx)}{\alpha} \in M_1(\Sigma)$. Thus $\mathcal{E}_\nu$ is non-empty and

$$\begin{aligned} & H(\rho|\rho_1 \otimes \rho_2) + I^W(\rho_1) + I^X(\rho_2) \\ = \quad & (1-\alpha)H(\nu|\rho_2) + \alpha H\left(\frac{\rho_2 - (1-\alpha)\nu}{\alpha}|\rho_2\right) I^X(\rho_2) \\ \geq \quad & \inf_{\rho_2 \in \mathcal{E}_\nu} \mathcal{U}(\nu, \rho_2) \end{aligned}$$

If $\alpha = 0$ then $\gamma = \delta_1$ and for every $\nu \in M_1(\Sigma)$ there is only one $\rho \in \mathcal{M}_1^1(\mathbb{R}_+ \times \Sigma)$ such that $F(\rho) = \nu$ which is $\rho = \delta_1 \otimes \nu$ and the announced result immediately follows. □

7.5 The k-blocks bootstrap

7.5.1 Proof of Theorem 2.5

Let us apply Corollary 2.1 to $\mathcal{L}^n = \frac{1}{n}\sum_{i=1}^n W_i^n \delta_{(x_i^n,\dots,x_{i+k-1}^n)}$. We obtain the desired result using the contraction principle on the map

$$\begin{array}{cccc} \mathcal{H}: & M_1(\Sigma^k) & \to & M_1(\Sigma) \\ & \rho^{(k)} & \mapsto & \mathcal{H}(\rho^{(k)}) = \frac{1}{k}\sum_{i=1}^k \rho_i^{(k)} \end{array}$$

which is continuous because it is Lipschitz. □

7.5.2 Proof of Corollary 2.7

Let $\nu \in M_1(\Sigma)$. Since $\mathcal{H}(\nu^{\otimes k}) = \nu$ and in this particular case $\mu^{(k)} = \mu^{\otimes k}$ we get

$$\begin{aligned} \widetilde{\mathcal{K}}(\nu; \mu^{(k)}) \quad & \leq \quad \frac{1}{k} H(\nu^{\otimes k}|\mu^{\otimes k}) \\ & = \quad H(\nu|\mu). \end{aligned}$$

On the other hand for every $\rho^{(k)}$ such that $\mathcal{H}(\rho^{(k)}) = \nu$ we have

$$H(\nu|\mu) = H(\frac{1}{k}\sum_{i=1}^k \rho_i^{(k)}|\mu) \leq \frac{1}{k}\sum_{i=1}^k H(\rho_i^{(k)}|\mu)$$

since $H(\cdot|\mu)$ is convex. To conclude the proof just notice that

$$\begin{aligned} H(\rho^{(k)}|\mu^{\otimes k}) \quad & = \quad H(\rho^{(k)}| \otimes_{i=1}^k \rho_i^{(k)}) + H(\otimes_{i=1}^k \rho_i^{(k)}|\mu^{\otimes k}) \\ & \geq \quad H(\otimes_{i=1}^k \rho_i^{(k)}|\mu^{\otimes k}) \\ & = \quad \sum_{i=1}^k H(\rho_i^{(k)}|\mu). \end{aligned}$$

□

A Topological properties of $(\mathcal{M}_1^1(\mathbb{R}_+ \times \Sigma), \Delta)$ and $(M_1^1(\mathbb{R}_+), W_1)$

A.1 $(\mathcal{M}_1^1(\mathbb{R}_+ \times \Sigma), \Delta)$ is a Polish space

A.1.1 $(\mathcal{M}_1^1(\mathbb{R}_+ \times \Sigma), \Delta)$ is complete

Let $(\rho^n)_{n\geq 1}$ be a Cauchy sequence of elements of $(\mathcal{M}_1^1(\mathbb{R}_+ \times \Sigma), \Delta)$. In particular $(\rho_1^n)_{n\geq 1}$ is a Cauchy sequence of elements of $(M_1^1(\mathbb{R}_+), W_1)$ which is a complete space (see e.g. [4]) so there exists a $\rho_1 \in M_1^1(\mathbb{R}_+)$ such that $W_1(\rho_1^n, \rho_1) \to 0$. Furthermore $(\rho^n)_{n\geq 1}$ is a Cauchy sequence of elements of $(M_1(\mathbb{R}_+ \times \Sigma), \beta_{BL,\delta})$ hence there exists a $\gamma \in M_1(\mathbb{R}_+ \times \Sigma)$ such that $\beta_{BL,\delta}(\rho^n, \gamma) \to 0$. Necessarily we have $\gamma_1 = \rho_1$ hence $\Delta(\rho^n, \gamma) \to 0$.

A.1.2 $(\mathcal{M}_1^1(\mathbb{R}_+ \times \Sigma), \Delta)$ is separable

Let R and E be dense countable subsets of $\mathbb{R}_+$ and Σ respectively. It is sufficient to prove that

$$\bigcup_{n\geq 1} \left\{ \frac{1}{n} \sum_{i=1}^n \delta_{(u_i^n, v_i^n)} \ : \ \text{for every } 1 \leq i \leq n,\ u_i^n \in R \text{ and } v_i^n \in E \right\}$$

is dense in $\mathcal{N}_1^1(\mathbb{R}_+ \times \Sigma)$ endowed with the distance Δ. Indeed, since $\mathcal{M}_1^1(\mathbb{R}_+ \times \Sigma)$ is a closed subset of $\mathcal{N}_1^1(\mathbb{R}_+ \times \Sigma)$ the announced claim follows. So let $\rho \in \mathcal{N}_1^1(\mathbb{R}_+ \times \Sigma)$. In particular $\int_{\mathbb{R}_+} x\rho_1(dx) < \infty$. Lets us denote by $(X_1, Y_1), \ldots, (X_n, Y_n), \ldots$ a sequence of independent random variables with common distribution ρ. According to Varadarajan's Lemma

$$\frac{1}{n} \sum_{i=1}^n \delta_{(X_i^n, Y_i^n)} \xrightarrow{w} \rho \ \text{ almost surely} \tag{A.1}$$

and according to the strong law of Large Numbers

$$\frac{1}{n} \sum_{i=1}^n X_i^n \to \int_{\mathbb{R}_+} x\rho_1(dx) \ \text{ almost surely} \tag{A.2}$$

So there exists a family $((x_i^n, y_i^n)_{1\leq i\leq n})_{n\geq 1}$ of elements of $\mathbb{R}_+ \times \Sigma$ such that

$$\frac{1}{n} \sum_{i=1}^n \delta_{(x_i^n, y_i^n)} \xrightarrow{w} \rho \ \text{ and } \quad \frac{1}{n} \sum_{i=1}^n X_i^n \to \int_{\mathbb{R}_+} x\rho_1(dx).$$

Since $R \times E$ is dense in $\mathbb{R}_+ \times \Sigma$ for every $n \geq 1$ and every $1 \leq i \leq n$ there exists $(u_i^n, v_i^n) \in R \times E$ such that $\max\{|u_i^n - x_i^n|, d_\Sigma(v_i^n, y_i^n)\} \leq 2^{-n}$. Clearly,

$$W_{1,\tilde{d}_{2,+}}\left(\frac{1}{n} \sum_{i=1}^n \delta_{(u_i^n, v_i^n)}, \frac{1}{n} \sum_{i=1}^n \delta_{(x_i^n, y_i^n)}\right) \leq 2^{-n}$$

hence $W_{1,\tilde{d}_{2,+}}(\frac{1}{n}\sum_{i=1}^{n}\delta_{(u_i^n,v_i^n)},\rho) \to 0$. Since both $\beta_{BL,\delta}$ and $W_{1,\tilde{d}_{2,+}}$ metrize the weak convergence topology we have that

$$\beta_{BL,\delta}(\frac{1}{n}\sum_{i=1}^{n}\delta_{(u_i^n,v_i^n)},\rho) \to 0.$$

Furthermore we have $\frac{1}{n}\sum_{i=1}^{n}\delta_{u_i^n} \stackrel{w}{\to} \rho_1$ and since

$$\begin{aligned} \left|\frac{1}{n}\sum_{i=1}^{n}u_i^n - \frac{1}{n}\sum_{i=1}^{n}x_i^n\right| &\leq \frac{1}{n}\sum_{i=1}^{n}|u_i^n - x_i^n| \\ &\leq 2^{-n} \end{aligned}$$

we obtain that $\frac{1}{n}\sum_{i=1}^{n}u_i^n \to \int_{\mathbb{R}_+} x\rho_1(dx)$ which together with the latter weak convergence ensure that $\frac{1}{n}\sum_{i=1}^{n}\delta_{u_i^n} \stackrel{W_1}{\to} \rho_1$, according to Theorem 7.11 in [26].

A.2 An approximation result on $(M_1^1(\mathbb{R}_+), W_1)$.

Lemma A.1. *For every $\rho \in M_1^1(\mathbb{R}_+)$ there exists for every $n \geq 1$ an elements ρ^n of*

$$\mathcal{A}_n = \left\{\nu \in M_1^1(\mathbb{R}_+),\ \exists\ (x_1,\ldots,x_n) \in (\mathbb{R}_+)^n, \nu = \frac{1}{n}\sum_{i=1}^{n}\delta_{x_n}\right\}$$

such that $\nu^n \stackrel{W_1}{\to} \rho$.

Proof By the same kind of argument as in the separability proof above one can construct a sequence $(\frac{1}{n}\sum_{i=1}^{n}\delta_{u_i^n})_{n\geq 1}$ such that $\frac{1}{n}\sum_{i=1}^{n}\delta_{u_i^n} \stackrel{W_1}{\to} \rho$. In particular $\frac{1}{n}\sum_{i=1}^{n}u_i^n \to 1$. So we only need to modify the u_i's in such a way that for every $n \geq 1$ their total sum equals n. For a fixed n we have three possibilities:
• If $\sum_{i=1}^{n}u_i = n$ we take $w_i^n = u_i$ for every $1 \leq i \leq n$.
• If $\sum_{i=1}^{n}u_i > n$ we look at the u_i's as the occupation masses of n cells by a mass u_i each. We pick uniformly at random the excess of mass until we get new occupation masses $w_1^n,\ldots,w_n$ such that $\sum_{i=1}^{n}w_i^n = n$.
• If $\sum_{i=1}^{n}u_i < n$ again we look at the u_i's as the occupation masses and add mass uniformly at random into the n cells until they contain a total mass of n. We call $w_1^n,\ldots,w_n$ the final occupation masses.

In all the cases considered above we have

$$\begin{aligned}
W_1(\frac{1}{n}\sum_{i=1}^n \delta_{u_i^n}, \frac{1}{n}\sum_{i=1}^n \delta_{w_i^n}) &= \sup_{\substack{f\in C_b(\mathbb{R}_+)\\ ||f||_L\leq 1}} \left\{ \left| \frac{1}{n}\sum_{i=1}^n f(u_i^n) - \frac{1}{n}\sum_{i=1}^n f(w_i^n) \right| \right\} \\
&\leq \sup_{\substack{f\in C_b(\mathbb{R}_+)\\ ||f||_L\leq 1}} \frac{1}{n}\sum_{i=1}^n |f(u_i^n) - f(w_i^n)| \\
&\leq \frac{1}{n}\sum_{i=1}^n |u_i^n - w_i^n| \\
&= \left| \frac{1}{n}\sum_{i=1}^n u_i^n - \frac{1}{n}\sum_{i=1}^n w_i^n \right| \\
&= \left| \frac{1}{n}\sum_{i=1}^n u_i^n - 1 \right| \to 0.
\end{aligned}$$

□

References

[1] P. Barbe and P. Bertail. *The weighted bootstrap*, volume 98 of *Lecture Notes in Statistics*. Springer-Verlag, New York, 1995.

[2] J. R. Baxter and N. C. Jain. A comparison principle for large deviations. *Proc. Amer. Math. Soc.*, 103(4):1235–1240, 1988.

[3] P. Billingsley. *Convergence of probability measures*. John Wiley & Sons Inc., New York, 1968.

[4] F. Bolley. Separability and completeness for the Wasserstein distance. In *Séminaire de probabilités XLI*, volume 1934 of *Lecture Notes in Math.*, pages 371–377. Springer, Berlin, 2008.

[5] E. Bolthausen. Markov process large deviations in τ-topology. *Stochastic Process. Appl.*, 25(1):95–108, 1987.

[6] N. R. Chaganty and R. L. Karandikar. Some properties of the Kullback-Leibler number. *Sankhyā Ser. A*, 58(1):69–80, 1996.

[7] A. Dembo and O. Zeitouni. Large deviations for subsampling from individual sequences. *Statist. Probab. Lett.*, 27(3):201–205, 1996.

[8] A. Dembo and O. Zeitouni. *Large deviations techniques and applications*, volume 38 of *Applications of Mathematics (New York)*. Springer-Verlag, New York, second edition, 1998.

[9] I. H. Dinwoodie and S. L. Zabell. Large deviations for exchangeable random vectors. *Ann. Probab.*, 20(3):1147–1166, 1992.

[10] M. D. Donsker and S. R. S. Varadhan. Asymptotic evaluation of certain Markov process expectations for large time. III. *Comm. Pure Appl. Math.*, 29(4):389–461, 1976.

[11] R. M. Dudley. *Real analysis and probability*, volume 74 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 2002. Revised reprint of the 1989 original.

[12] P. Dupuis and R. S. Ellis. *A weak convergence approach to the theory of large deviations*. Wiley Series in Probability and Statistics: Probability and Statistics. John Wiley & Sons Inc., New York, 1997. A Wiley-Interscience Publication.

[13] B. Efron. *The jackknife, the bootstrap and other resampling plans*. CBMS-NSF Regional Conference Series in Applied Mathemati, SIAM, Philadelphia, Pa., 1982.

[14] P. Eichelsbacher and U. Schmock. Large deviations of U-empirical measures in strong topologies and applications. *Ann. Inst. H. Poincaré Probab. Statist.*, 38(5):779–797, 2002.

[15] R. S. Ellis, J. Gough, and J. V. Pulé. The large deviation principle for measures with random weights. *Rev. Math. Phys.*, 5(4):659–692, 1993.

[16] E. Giné. Lectures on some aspects of the bootstrap. In *Lectures on probability theory and statistics (Saint-Flour, 1996)*, volume 1665 of *Lecture Notes in Math.*, pages 37–151. Springer, Berlin, 1997.

[17] M. Grunwald. Sanov results for Glauber spin-glass dynamics. *Probab. Theory Related Fields*, 106(2):187–232, 1996.

[18] P. Hall. *The bootstrap and Edgeworth expansion*. Springer Series in Statistics. Springer-Verlag, New York, 1992.

[19] P. Hall and E. Mammen. On general resampling algorithms and their performance in distribution estimation. *Ann. Statist.*, 22(4):2011–2030, 1994.

[20] S. Holmes and G. Reinert. Stein's method for the bootstrap. In *Stein's method: expository lectures and applications*, volume 46 of *IMS Lecture Notes Monogr. Ser.*, pages 95–136. Inst. Math. Statist., Beachwood, OH, 2004.

[21] S. Kullback. *Information theory and statistics*. Dover Publications Inc., Mineola, NY, 1997. Reprint of the second (1968) edition.

[22] C. L. Liu. *Introduction to combinatorial mathematics*. McGraw-Hill Book Co., New York, 1968.

[23] J. Najim. A Cramér type theorem for weighted random variables. *Electron. J. Probab.*, 7:no. 4, 32 pp. (electronic), 2002.

[24] J. Præstgaard and J. A. Wellner. Exchangeably weighted bootstraps of the general empirical process. *Ann. Probab.*, 21(4):2053–2086, 1993.

[25] J. Trashorras. Large deviations for symmetrised empirical measures. *J. Theoret. Probab.*, 21(2):397–412, 2008.

[26] C. Villani. *Topics in optimal transportation*, volume 58 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 2003.

[27] R. Wang, X. Wang, and L. Wu. Sanov's theorem in the Wasserstein distance: a necessary and sufficient condition. *Statist. Probab. Lett.*, 80(5-6):505–512, 2010.

[28] A. T. A. Wood. Bootstrap relative errors and sub-exponential distributions. *Bernoulli*, 6(5):809–834, 2000.